



# Risk with random normalizing factors in the white gaussian noise additive model

Fabien Chiabrando

## ► To cite this version:

Fabien Chiabrando. Risk with random normalizing factors in the white gaussian noise additive model. 2007. hal-00174530

**HAL Id: hal-00174530**

**<https://hal.science/hal-00174530>**

Preprint submitted on 24 Sep 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RISK WITH RANDOM NORMALIZING FACTORS IN THE WHITE GAUSSIAN NOISE ADDITIVE MODEL

F.CHIABRANDO

**ABSTRACT.** In the context of minimax theory we develop a new approach based on pretesting. The first step of this approach consists in testing some structural assumption imposed on the underlying function. According to the result we use a relevant estimation procedure that allows to improve significantly the quality of estimation. We apply this general set-up to the estimation of an unknown multidimensionnal signal in the White Gaussian Noise model. The structure we test here is the additivity hypothesis. The mathematical description of this approach leads to the notion of random (depending on datas) rate of estimation. Under some additional assumption our construction leads to adaptive estimator w.r.t. rate of convergence.

## 1. INTRODUCTION

**1.1. Minimax approach.** This paper is devoted to the estimation of the signal observed in multivariate gaussian white noise model.

$$(1.1) \quad dX_\varepsilon(t) = f(t)dt + \varepsilon dB(t), \quad \forall t \in [0, 1]^d,$$

where  $B(\cdot)$  is the standard Brownian sheet and  $0 < \varepsilon < 1$  is the noise level.

To measure performance of estimators statisticians often use the minimax quadratic risk determined by  $L_2$ -norm on  $[0, 1]^d$ . For given set  $\Sigma$  subset of  $L_2([0, 1]^d)$  we define a maximal risk

$$(1.2) \quad R_\varepsilon(\hat{f}_\varepsilon, \Sigma) = \sup_{f \in \Sigma} E_f \|\hat{f}_\varepsilon - f\|_2^2,$$

and we interesting in the asymptotics (minimax rate of convergence) of the minimax risk

$$(1.3) \quad \varphi_\varepsilon(\Sigma)^2 \asymp \inf_{\hat{f}} R_\varepsilon(\hat{f}_\varepsilon, \Sigma),$$

where the inf is taken over all estimators. The main difficulty arises in estimation of a multivariate function is the so called curse of dimensionality : the larger dimension , the worse quality of estimation. In particular the dimension effects on the MRC. For example, if  $\Sigma = \Sigma_d(\beta, L)$  , where  $\Sigma_d(\beta, L)$  isotropic Holder or Sobolev ball with smoothness parameter  $\beta$ , then [16]

$$(1.4) \quad \varphi_\varepsilon(\beta, d) = \varepsilon^{\frac{2\beta}{2\beta+d}}.$$

As we see, the higher d is, the lower is the minimax rate. Consequently the asymptotical result becomes irrelevant because it is applicable only for unreasonably small noise level  $\varepsilon > 0$ . This problem arises because isotropic Holder or Sobolev balls are too massive. A way to overcome the curse of dimensionality is to use the poor functional classes for description of the model. Typically, the "poverty" of a functional class can be expressed in terms of restrictions on its metric entropy and there are several possibilities to do that. One can suppose, for instance, that regularity  $\beta$  grows proportionnaly with dimension. In this case the dimension will disappear from the expression of the MRC. However such assumption is quite unrealistic and restrictive and it can lead to unadequate modeling.

Another way consists in imposing an additional structure on the signal. In particular, Stone [30] proposed to use the following structure, called additive,

$$(1.5) \quad f(t) = f_1(t_1) + \dots + f_d(t_d),$$

when  $f_i \in \Sigma_1(\beta, L)$  where  $\Sigma_1(\beta, L)$  is either Holder or Sobolev ball of univariate functions. He showed that the minimax rate of convergence under additivity hypothesis is given by

$$(1.6) \quad \varphi_\varepsilon(\beta) = \varphi_\varepsilon(\beta, 1) = \varepsilon^{\frac{2\beta}{2\beta+1}},$$

i.e. the quality of estimation does not depend on the dimension.

It also worth mentioning that there exists also a vast literature on estimation of components  $f_i$  when (1.5) holds. For example, algorithmic methods as 'backfitting' introduced by Breiman and Friedman [3], or non-iterative method as 'marginal estimation' proposed in [26]. Consistency and asymptotic properties were studied over these two approaches see [27], [29].

Other structures which have been studied in recent literature are

- **Single-index model** : Let  $e$  in  $\mathbb{R}^d$ , assume that  $f(x) = F(e^T x)$  for univariate  $F$ .
- **Projection pursuit model** : Let  $e_1, \dots, e_d$  in  $\mathbb{R}^d$  and assume that there are univariate function  $f_i$  such that  $f(x) = \sum_{i=1}^d f_i(e_i^T x)$
- **Multi-index model** : Let  $e_1, \dots, e_m$  in  $\mathbb{R}^d$  with  $m < d$  and assume that  $f(x) = F(e_1^T x, \dots, e_m^T x)$  with  $m$ -dimensional  $F$ .

The existing results show that the use of structural models usually leads to the improvement of the quality of estimation w.r.t the rate  $\varphi_\varepsilon(\beta, d)$ , Chen [6], Golubev[11].

Thus, the structural assumption usually improves the accuracy of estimation. However they are very strong and can lead to inadequate model. Indeed, it is quite restrictive to suppose that the underlying function possesses additive or single-index structure in the whole domain of observation. To avoid this problem and in the same time to guarantee the flexibility of the modeling one can use so called adaptive approach.

**1.2. Adaptive approach.** For the model discussed above, this approach consists in adapting an estimation procedure to eventual structure of the signal. Therefore, we hope to improve accuracy of estimation as soon as the underlying function belong to  $\Sigma_0$  a "poor" subset of  $\Sigma$  i.e.  $\varphi_\varepsilon(\Sigma_0) \ll \varphi_\varepsilon(\Sigma)$ . Thus we seek an estimator  $f_\varepsilon^{(a)}$ , called adaptive, such that

$$(1.7) \quad \limsup_{\varepsilon \rightarrow 0} \varphi_\varepsilon(\Sigma)^{-2} R_\varepsilon(f_\varepsilon^{(a)}, \Sigma) < \infty,$$

$$(1.8) \quad \limsup_{\varepsilon \rightarrow 0} \varphi_\varepsilon(\Sigma_0)^{-2} R_\varepsilon(f_\varepsilon^{(a)}, \Sigma_0) < \infty.$$

The existence as well as non existence of adaptive estimators was extensively studied during last two decades, [1] [7] [8], [9], [10], [12], [14], [15], [17], [18],[19], [23], [24], [4] and [25] among others. In our model, we present an estimator that satisfies to (1.7) and (1.8) when  $\Sigma_0$  corresponds to the subset of additive functions (1.5) belonging to isotropic Sobolev ball.

Under assumption that an adaptive estimator have been built, one may also wonder what kind of informations the estimation procedure brings to the statistician? Unfortunately even having constructed an estimator satisfying to (1.7) and (1.8), one can say nothing on its accuracy of estimation. Indeed putting

$$(1.9) \quad \psi_\varepsilon(f) = \begin{cases} \varphi_\varepsilon(\Sigma_0), & \text{if } f \in \Sigma_0, \\ \varphi_\varepsilon(\Sigma), & \text{if } f \in \Sigma \setminus \Sigma_0, \end{cases}$$

(1.7) and (1.8) are equivalent to

$$(1.10) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{f \in \Sigma} E_f \left\{ \psi_\varepsilon^{-1}(f) \|f_\varepsilon^{(a)} - f\|_2 \right\}^2 = C < \infty.$$

As we see accuracy  $\psi_\varepsilon(f)$  depends on the estimated function which is unknown. More exactly, it depends on whether  $f$  belongs to  $\Sigma_0$  or not. This information can not be obtained from noisy datas. This impossibility of computing accuracy of an adaptive estimator is an unavoidable payment for the adaptive property. For

example, this lack of information keeps the statistician from deducing adaptive confidence ball for  $f$  from its estimation. Assume that  $\Sigma = \Sigma_d(\beta, L)$  is a  $d$ -dimensional isotropic Sobolev ball and  $\Sigma_0$  is the subset of additive functions i.e. that satisfy to (1.5). Then referring to (1.4) and (1.6) one has that

$$(1.11) \quad \psi_\varepsilon(f) = \begin{cases} \varepsilon^{\frac{2\beta}{2\beta+1}}, & \text{if } f \in \Sigma_0, \\ \varepsilon^{\frac{2\beta}{2\beta+d}}, & \text{if } f \in \Sigma \setminus \Sigma_0. \end{cases}$$

Let us suppose that some adaptive estimator  $f_\varepsilon^*$  w.r.t. the family  $\{\Sigma, \Sigma_0\}$  is built. Using Markov inequality we get for any  $0 < \nu < 1$  and any  $f \in \Sigma$

$$(1.12) \quad P_f \left( \|f - f_\varepsilon^*\| \leq \sqrt{\frac{C}{\nu}} \psi_\varepsilon(f) \right) \geq 1 - \nu.$$

It means that with probability equal to  $1 - \nu$ ,  $f$  belongs to the ball with center  $f_\varepsilon^*$  and with radius of order  $\psi_\varepsilon(f)$ . In accordance whether the underlying function is truly additive or not, the radius of this 'artificial' confidence ball would be of the order either  $\varepsilon^{\frac{2\beta}{2\beta+1}}$  or  $\varepsilon^{\frac{2\beta}{2\beta+d}}$ , which both are optimal in minimax sense. We call it artificial because it is not really confidence ball, since its radius depends on the unknown function. One way to avoid this difficulty may consist in replacing  $\psi_\varepsilon(f)$  in (1.12) by some random variable computed from the datas. To realize this idea, a new definition of minimax risk is required. Lepski in [21] proposes to measure accuracy of an estimator  $\hat{f}_\varepsilon$  using

$$(1.13) \quad R_\varepsilon^{(r)}(\hat{f}_\varepsilon, \Sigma, \rho_\varepsilon) = \sup_{f \in \Sigma} E_f \left[ \rho_\varepsilon^{-1} \|\hat{f}_\varepsilon - f\|_2 \right]^2,$$

where  $\rho_\varepsilon$  is some random variable obtained from datas. This kind of risk is called risk with random normalizing factor (RNF). The issue of this approach is a couple  $(\rho_\varepsilon, \hat{f}_\varepsilon)$  such that  $R_\varepsilon^{(r)}(\hat{f}_\varepsilon, \Sigma, \rho_\varepsilon)$  is bounded by some constant  $C$ . Then for any  $0 < \nu < 1$  and any  $f \in \Sigma$

$$(1.14) \quad P_f \left( \|f - \hat{f}_\varepsilon\| \leq \sqrt{\frac{C}{\nu}} \rho_\varepsilon \right) \geq 1 - \nu.$$

Thus, contrary to (1.12), (1.14) provides a 'real' confidence ball for the observed signal. Since its radius now is computable. However this new approach requires to be able to compare different couples  $(\tilde{\rho}_\varepsilon, \tilde{f}_\varepsilon)$ . Indeed, the randomness of considered normalizing factors removes the natural order used in minimax theory. We present the exact definitions concerning RNF in the next section.

This approach has been already applied to the models where eventual structure can improve the accuracy of estimation. First, let us mention [21] where the parametric and regularity hypothesis were considered in context of estimation of univariate function. Next two papers [22] and [32] deal with estimation of multivariate functions. For WGN model the estimation procedure based on testing of dimensionality hypothesis ( $\exists i_1 < \dots < i_s$  in  $\{1, \dots, d\} : f(x) = F(x_{i_1}, \dots, x_{i_s})$ ) was proposed in [22]. In [32], the estimator of the density ( $f(x), x \in \mathbb{R}^d$ ) which is based on testing of independance hypothesis ( $f(x) = \prod_{i=1}^d f_i(x_i)$ ) was solved. Both papers propose the construction of optimal RNF in the sense of Lepski and present adaptive estimator w.r.t. the tested structure.

We will apply this approach in the case where  $\Sigma$  is isotropic ellipsoid with polynomial decreasing axes and  $\Sigma_0 \subset \Sigma$  consists in additive functions. For this problem, the construction of optimal RNF and efficient estimator is based on the hypothesis testing of additive structure. We would like to stress that under additional assumptions the corresponding estimation procedure has adaptive properties. More precisely, having required that the first type error probability of the test tends to zero fast enough, we arrive to an adaptive estimator (1.10).

The paper is organized as follows. In section 2 we discuss in detail the minimax risk with RNF. The section 3 is devoted to the construction of RNF and the estimator which are based on testing the additivity hypothesis as well as the presentation of main results. In section 4, we discuss the way how to adapt the

results obtained in the framework of WGN to the regression model with deterministic design. The proofs are postponed to the last part of the paper.

## 2. RISK WITH RANDOM NORMALIZING FACTOR

Let us denote by  $\mathfrak{E}_\varepsilon$  the set of bounded RNF  $\rho_\varepsilon$  i.e. random variables measurable w.r.t.  $X_\varepsilon(\cdot)$  and taking two values  $\{\varphi_\varepsilon(\Sigma), b_\varepsilon\}$ ,  $0 < b_\varepsilon < \varphi_\varepsilon(\Sigma)$ . By  $\mathfrak{A}_\varepsilon$  we denote the set of all possible estimators.

For any  $\rho_\varepsilon \in \mathfrak{E}_\varepsilon$  and  $\hat{f}_\varepsilon(\cdot) \in \mathfrak{A}_\varepsilon$  we defined the risk

$$(2.1) \quad R_\varepsilon^{(r)}(\hat{f}_\varepsilon, \Sigma, \rho_\varepsilon) = \sup_{f \in \Sigma} E_f^\varepsilon \left\{ \rho_\varepsilon^{-1} \|\hat{f}_\varepsilon - f\|_2 \right\}^2.$$

The reason to introduce the risk (2.1) consists in the following. We understand the 'improvement' of accuracy as the fulfillment of the event  $\{\rho_\varepsilon = b_\varepsilon\}$ . As we will see the fact whether this event holds or not depends on the acceptance or rejection of the hypothesis

$$\mathbf{H}_0 : f(\cdot) \in \Sigma_0.$$

Clearly we would like that  $\{\rho_\varepsilon = b_\varepsilon\}$  holds with prescribed probability at least if  $f$  belongs to  $\Sigma_0$ . To realize this idea we proceed as follows.

First let us note that  $\sup_{f \in \Sigma_0} P_f(\{\rho_\varepsilon \neq b_\varepsilon\})$  can be viewed as the first type error for the test of  $\mathbf{H}_0$ . This leads to the definition of the set of 'reasonable' RNF's.

**Definition 2.1.** For a given  $0 < \alpha_\varepsilon < 1$ ,

$$(2.2) \quad \Omega = \Omega(\alpha) = \left\{ \rho_\varepsilon \in \mathfrak{E}_\varepsilon : \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{f \in \Sigma_0} P_f^\varepsilon \{\rho_\varepsilon = \varphi_\varepsilon(\Sigma)\} \leq 1 \right\}.$$

Thus, the consideration of  $\rho_\varepsilon \in \Omega(\alpha)$  guarantees the 'improvement' w.r.t. to  $\varphi_\varepsilon(\Sigma)$  if  $f \in \Sigma_0$  with probability larger than  $1 - \alpha_\varepsilon$ . In this context, it seems that  $\alpha_\varepsilon$  should be chosen tending to zero as fast as possible. Actually it is not true because the value  $b_\varepsilon$  and  $\alpha_\varepsilon$  are related: the faster  $\alpha_\varepsilon$  tends to zero, the 'closer'  $b_\varepsilon$  to  $\varphi_\varepsilon(\Sigma)$ . Let also note that the introduction of the set  $\Omega$  allows us to propose a criterion of optimality [21] for the risk of type (2.1).

**Definition 2.2.** A RNF  $(\rho_\varepsilon^*)_\varepsilon = \{\varphi_\varepsilon(\Sigma), b_\varepsilon^*\} \in \Omega$  is called  $\alpha$ -optimal w.r.t. the family  $\{\Sigma, \Sigma_0\}$  when :

(1) There exists an estimator  $f_\varepsilon^*$  such that

$$\limsup_{\varepsilon} R_\varepsilon^{(r)}(f_\varepsilon^*, \Sigma, \rho_\varepsilon^*) \leq C.$$

(2) If there exist  $\tilde{\rho}_\varepsilon = \{\varphi_\varepsilon(\Sigma), \tilde{b}_\varepsilon\} \in \Omega$  such that  $\lim_{\varepsilon} \frac{\tilde{b}_\varepsilon}{b_\varepsilon^*} = 0$ ,

$$\text{then} \quad \liminf_{\varepsilon} \inf_{f_\varepsilon} R_\varepsilon^{(r)}(f_\varepsilon, \Sigma, \tilde{\rho}_\varepsilon) = \infty.$$

Thus for a given  $\alpha_\varepsilon$ , we compare RNF's from  $\Omega(\alpha)$  thanks to their second term  $b_\varepsilon$ . Indeed this term can be viewed as an improvement w.r.t. to minimax rate of convergence on  $\Sigma$ . On the other hand, under rather vast assumptions  $b_\varepsilon \geq \varphi_\varepsilon(\Sigma_0)$  [21]. Clearly the closer  $b_\varepsilon$  to  $\varphi_\varepsilon(\Sigma_0)$  the best improvement is.

**Definition 2.3.** Let  $(\rho_\varepsilon^*)_\varepsilon$  be an  $\alpha$ -optimal w.r.t.  $\{\Sigma, \Sigma_0\}$ . Then any estimator satisfying to (2.3) is called  $\alpha$ -adaptive.

Due to definition 2.2 and 2.3, the issue of this approach consists in finding a couple  $(\rho_\varepsilon^*, f_\varepsilon^*)$  such that the risk (2.1) is controlled.

**Remark 2.1.** By definition of RNF  $\rho_\varepsilon \leq \varphi_\varepsilon(\Sigma)$  for all  $\varepsilon \in (0, 1)$ , therefore any  $\alpha$ -adaptive estimator is also minimax on the  $\Sigma$ .

**Remark 2.2.** Typically an  $\alpha$ -estimator is constructed on the following way

$$(2.3) \quad f_\varepsilon^* = \begin{cases} \widehat{f}_\varepsilon^{(0)}(\cdot), & \text{if } \mathbf{H}_0 \text{ is accepted,} \\ \widehat{f}_\varepsilon(\cdot), & \text{if not,} \end{cases}$$

where  $\widehat{f}_\varepsilon, \widehat{f}_\varepsilon^{(0)}$  are minimax estimators on the sets  $\Sigma$  and  $\Sigma_0$ . In this case,  $\alpha_\varepsilon$  can be viewed as an upper bound for the first type error of the test of  $\mathbf{H}_0$ .

**Remark 2.3.** Lepski proved in [21] that if  $\alpha_\varepsilon = \mathcal{O}(\varphi_\varepsilon(\Sigma_0)^2)$  then  $f_\varepsilon^*$  satisfies to (2.3) is also adaptive in the sense that (1.10) holds.

**Remark 2.4.** We present the definition of  $\alpha$ -optimal RNF and  $\alpha$ -adaptive estimator only in the case of a single hypothesis. Let us mention that this definition is extended [22] to arbitrary many hypothesis which allows, in particular to adapt simultaneously to different structures.

### 3. MODEL, CONSTRUCTION AND MAIN RESULT

**3.1. Model.** Let statistical experiment be generated by the observation  $X^\varepsilon$  which is the sample of the stochastic process  $X_\varepsilon(\cdot)$  satisfying on the d-cube  $[0, 1]^d$  the stochastic differential equation

$$(3.1) \quad dX_\varepsilon(t) = f(t).dt + \varepsilon dB(t). \quad \forall t \in [0, 1]^d,$$

where  $B(\cdot)$  is standard Wiener process and  $\varepsilon$  is the level of the noise.

Though, we would prefer to look to this model under an equivalent notation. Let  $(\underline{\phi}_k, k \in \mathbb{N})$  be an orthonormal basis of  $L_2([0, 1])$  such that

$$(3.2) \quad \int_{[0,1]} \underline{\phi}_k(\underline{t}).dt = \delta_{0,k},$$

where  $\delta$  is the Kronecker symbol. For instance one can use  $\underline{\phi}_0 = 1$  and for  $k \neq 0$ ,

$$(3.3) \quad \underline{\phi}_{2k}(\underline{t}) = \sqrt{2} \cos(2\pi k \underline{t}), \quad \underline{\phi}_{2k+1}(\underline{t}) = \sqrt{2} \sin(2\pi k \underline{t}).$$

Obviously other choices are possible. For a multinindex  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$ ,

$$(3.4) \quad \phi_k(t) = \phi_k(\underline{t}_1, \dots, \underline{t}_d) = \underline{\phi}_{k_1}(\underline{t}_1) \dots \underline{\phi}_{k_d}(\underline{t}_d),$$

that define an orthormal basis of  $L_2([0, 1]^d)$ . Then for any  $f \in L_2([0, 1]^d)$ , one can use the following  $L_2$ -expansion :

$$f = \sum_{k \in \mathbb{N}^d} \theta_k \phi_k,$$

where

$$(3.5) \quad \theta_k = \theta_k(f) = \int_{[0,1]^d} f(t) \phi_k(t) dt.$$

Thus, we consider that new observations are  $(\theta_k(f))_{k \in \mathbb{N}^d}$  and then (3.1) is equivalent to :

$$(3.6) \quad y_k = \theta_k + \varepsilon \zeta_k, \quad \forall k \in \mathbb{N}^d,$$

when  $y_k = \int_{[0,1]^d} \phi_k(t) dX_\varepsilon(t)$  and  $\zeta_k = \int_{[0,1]^d} \phi_k(t) dB(t)$ . Owing to the orthormality of the basis  $\phi_k$ ,  $\zeta_k$  are i.i.d. standard gaussian variables. Moreover because of the equivalence between (3.1) and (3.6), we identify in the following the observed function  $f$  and its expansion  $\theta = (\theta_k)_{k \in \mathbb{N}^d}$ .

We focus on an isotropic Sobolev ellipsoid define with two parameters  $\beta > 0$  and  $L > 0$  by :

$$(3.7) \quad \Sigma = \Sigma(\beta, L) = \left\{ f \in L_2 : \sum_{k \in \mathbb{N}^d} \theta_k^2 \left( 1 + \sum_{j=1}^d k_j^{2\beta} \right) < L \right\},$$

and we define the subset of additive function

$$(3.8) \quad \Sigma_0 = \left\{ f \in \Sigma : \exists f_i \in \mathcal{F}(\mathbb{R}^d, \mathbb{R}) : \forall t = (t_i) \in [0, 1]^d, f(t) = \sum_{i=1}^d f_i(t_i) \right\}.$$

**Remark 3.1.** *These class of functions can be linked with usual isotropic Sobolev class defined as*

$$\underline{\Sigma}(\beta, L) = \left\{ f : [0, 1]^d \rightarrow \mathbb{R} : \|f\|^2 + \sum_{i=1}^d \left\| \frac{\partial^\beta}{\partial x_i^\beta} f \right\|^2 \leq L^2 \right\},$$

as soon as we assume that  $f$  is 1-periodic.

**3.2. Construction.** Let first discuss briefly the pre-testing step and the deduced estimation procedure. Based on Lepski's method, we use minimax projection estimators on  $\Sigma$  and  $\Sigma_0$  to compute an efficient estimator of  $d(f, \Sigma_0) = \inf_{g \in \Sigma_0} \|f - g\|$ . Then, accuracy of this estimation would impose us a threshold beyond we reject additive structure for  $f$ . That is so called decision rule. According to whether our test rejects or accepts the hypothesis of additive structure, we select either minimax projection estimator on  $\Sigma$  or else on  $\Sigma_0$  to estimate the underlying function. Remark that efficiency of the method is deeply related to the accuracy of estimation of  $d(f, \Sigma_0)$ .

Now we give notations necessary to our construction. Set the subsets of  $\mathbb{N}^d$  :

$$\begin{aligned} \Gamma_\varepsilon &= \{k \in \mathbb{N}^d : k_j < N_\varepsilon \ \forall j = 1 \dots d\} & \text{where} & \quad N_\varepsilon = \varepsilon^{-2/(2\beta+d)}, \\ \Delta_\varepsilon &= \{k \in \mathbb{N}^d : k_j < N_{0,\varepsilon} \ \forall j = 1 \dots d\} & \text{where} & \quad N_{0,\varepsilon} = 2 \left( \varepsilon^2 \sqrt{\ln(1/\alpha_\varepsilon)} \right)^{-2/(4\beta+d)}. \end{aligned}$$

For a given  $k \in \mathbb{N}^d$ , set

$$(3.9) \quad G_k = \{i : k_i \neq 0\},$$

then let us define the following multiindex sets

$$\begin{aligned} I_\varepsilon &= \Delta_\varepsilon \cap \{k \in \mathbb{N}^d : |G_k| \leq 1\}, \\ \Lambda_\varepsilon &= \Delta_\varepsilon \setminus I_\varepsilon. \end{aligned}$$

Construction of the test is based on the nullity of a large part of Fourier coefficients as soon as additive structure (1.5) holds. According to the orthogonality of the projection basis and condition (3.2), one have

$$(3.10) \quad [f \in \Sigma_0] \iff [|G_k| > 1 \Rightarrow \theta_k(f) = 0].$$

Let us recall that the minimax rate of convergence on  $\Sigma$  is  $\varepsilon^{\frac{2\beta}{2\beta+d}}$ , and Fourier expansion of an asymptotically minimax estimator is :

$$(3.11) \quad \hat{\theta}_{\varepsilon,k} = \begin{cases} y_k, & \text{if } k \in \Gamma_\varepsilon, \\ 0, & \text{if not.} \end{cases}$$

According to (3.10), one can easily check that the estimator which coefficients are

$$(3.12) \quad \hat{\theta}_{\varepsilon,k}^0 = \begin{cases} y_k, & \text{if } k \in I_\varepsilon, \\ 0, & \text{if not,} \end{cases}$$

attains the minimax rate of convergence (1.6) on  $\Sigma_0$ .

General adaptation procedure inspired by Lepski's method, leads us to look to the difference in  $L_2$  norm between minimax estimators of  $f$  on  $\Sigma$  and on  $\Sigma_0$  to detect additivity of the signal. The lower is this value, the higher the probability of a true additive structure for  $f$  is. Indeed

$$\sum_{k \in \mathbb{N}^d} \left( \hat{\theta}_{\varepsilon,k}^0 - \hat{\theta}_{\varepsilon,k} \right)^2 = \sum_{k \in \Lambda_\varepsilon} y_k^2 \approx \sum_{k \in \Lambda_\varepsilon} \theta_k^2 \approx \|f - p_{\Sigma_0}(f)\|^2 = d^2(f, \Sigma_0).$$

Such biased estimator can be modified to fit to  $\|f - p_{\Sigma_0}(f)\|^2$  in a better way i.e. with a lower bias. Consequently we interest in estimating  $\sum_{k \in \Lambda_\varepsilon} \theta_k(f)^2$  with the statistic

$$T_\varepsilon = \sum_{k \in \Lambda_\varepsilon} (y_k^2 - \varepsilon^2).$$

Then our decision rule is based on the acceptance of hypothesis  $\mathbf{H}_0$  when  $T_\varepsilon$  is sufficiently small and rejecting as soon as  $T_\varepsilon$  is larger than a determined threshold. Thus let us introduce the event

$$(3.13) \quad A_\varepsilon = \{T_\varepsilon \leq \lambda a_\varepsilon^2\},$$

when  $\lambda$  is a well-chosen constant that does not depend on the model and the threshold is

$$(3.14) \quad a_\varepsilon = (\varepsilon^2 \sqrt{\ln(1/\alpha_\varepsilon)})^{2\beta/(4\beta+d)}.$$

Let us postpone to the proof the explicit definition of constant  $\lambda$ . Finally we define the estimator  $\theta_\varepsilon^*$  and the RNF  $\rho_\varepsilon^*$  as follows

$$(3.15) \quad \rho_\varepsilon^* = \begin{cases} a_\varepsilon, & \text{on } A_\varepsilon, \\ \varphi_\varepsilon(\Sigma), & \text{if not,} \end{cases}$$

and

$$(3.16) \quad \theta_\varepsilon^* = \begin{cases} \widehat{\theta}_\varepsilon^0, & \text{on } A_\varepsilon, \\ \widehat{\theta}_\varepsilon, & \text{if not,} \end{cases}$$

**3.3. Results.** First of all, according to definition 2.2, any  $\alpha$ -optimal RNF must adapt to additive structure with prescribed probability error. Indeed error made in rejecting additive structure when it holds need to be controlled by  $\alpha_\varepsilon$ . Such assumption requires a sharp estimation of  $d(f, \Sigma_0)$ . The following lemma proves that under additivity,  $T_\varepsilon$  is a relevant estimator of  $d^2(f, \Sigma_0)$  and consequently that  $A_\varepsilon$  is a suitable decision rule w.r.t. to definition of  $\Omega$ .

**Lemma 3.1.** *The RNF  $(\rho_\varepsilon^*)_\varepsilon$  belongs to  $\Omega(\alpha)$ .*

Next we requires that  $\alpha_\varepsilon$  does not vanish too fast. Indeed let remark that the 'improvement' rate  $a_\varepsilon$  is slowed down by a term  $-\ln(\alpha_\varepsilon)$ . The following condition imposes that this term remains negligible w.r.t. any power of  $\varepsilon$  i.e. that  $\alpha_\varepsilon$  does not decrease exponentially fast to zero.

**Condition 1.** *The sequence  $(\alpha_\varepsilon)_\varepsilon$  satisfies to*

$$\exists a > 0 \quad : \quad \forall n \geq 1, \quad 1 > \alpha_\varepsilon > \varepsilon^a.$$

According to the previous discussion on the optimal choice of  $\alpha_\varepsilon$ , one can insure that such condition is not restrictive at all. Under condition above, we manage to state optimality of our procedure w.r.t. to definition 2.2.

**Theorem 3.1.** *Assume that  $\alpha_\varepsilon$  satisfies to condition 1, then the RNF  $(\rho_\varepsilon^*)_\varepsilon$  is  $\alpha$ -optimal w.r.t. the family  $\{\Sigma, \Sigma_0\}$  and  $\theta_\varepsilon^*$  is  $\alpha$ -adaptive.*

Using link between  $\alpha$ -adaptation and adaptation underlined above, one deduces that under condition on  $\alpha$ , estimator (3.16) is also adaptive in the common sense for the family  $\{\Sigma, \Sigma_0\}$ .

**Corollary 3.1.** *Let  $\alpha_\varepsilon = \mathcal{O}\left(\varepsilon^{\frac{4\beta}{2\beta+d}}\right)$  as  $\varepsilon \rightarrow 0$ , then  $\theta_\varepsilon^*$  is adaptive estimator i.e. satisfies to (1.10) with  $\psi_\varepsilon(f)$  define in (1.11).*

Let us comment these results. First, this procedure not only provides adaptive estimator (corollary 3.1) but also guarantees more precise coverage of the underlying function w.r.t. minimax accuracy, under acceptance of additive structure. This approach can be viewed as a trade-off between adaptation and minimax theories. First we have seen that under non restrictive conditions,  $\alpha$ -adaptive estimator keeps theoretical adaptative properties. Moreover the optimal RNF  $\rho_\varepsilon^*$  does not depend on the unknown  $f$  yet it largely improves minimax accuracy  $\varphi_\varepsilon(\Sigma)$  as soon as  $A_\varepsilon$  holds. However there exists an unavoidable payment



for computation of accuracy. Indeed, let notice that the best accuracy attained by the optimal RNF (3.15) is rather rougher than minimax rate on  $\Sigma_0$  (1.6), even when additive structure is accepted.

Moreover one may briefly discuss influence of the choice of  $\alpha_\varepsilon$ . As we see the lower is  $\alpha_\varepsilon$ , the poorer  $\alpha$ -adaptive accuracy is. For instance, if  $\alpha_\varepsilon$  tends to zero then the term  $\ln(1/\alpha_\varepsilon)$  asymptotically damages the accuracy  $a_\varepsilon$ . Moreover if statistician need to have an 'exponential' control of the first type error of the test i.e. impose  $\alpha_\varepsilon = e^{-\varepsilon^{-a}}$  then the accuracy of any  $\alpha$ -optimal RNF is largely reduced (depending on constant  $a > 0$ ). Though corollary 3.1 requires that  $\alpha_\varepsilon$  vanishes fast enough in order to deduce adaptive estimator. In conclusion, statistician may choose  $\alpha_\varepsilon$  one compared to what he first demand to the procedure : adaptive theoretical properties or the best improvement for the available accuracy of estimation.

#### 4. CONNECTION WITH THE REGRESSION MODEL

**4.1. Introduction.** The use of WGN model enables us to avoid superfluous technicalities not directly related to considered problem. However this model is rather unadapted to realistic statistical applications. One may wonder whether construction of an optimal RNF w.r.t. additive structure works in the same way for more realistic models or not. Below we discuss how to adapt our estimation procedure to the regression model.

First, let us consider the unidimensional regression model

$$(4.1) \quad Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

with  $\varepsilon_i$  are Gaussian zero-mean noise variables and  $f \in L_2(\mu)$  when  $\mu$  is a measure absolutely continuous w.r.t. Lebesgue measure, noted  $\lambda$ , on  $\mathbb{R}$ . In the following, we denote  $w(\cdot) = \frac{\partial \mu}{\partial \lambda}(\cdot)$ . In this section we will consider the regression model where the design  $(x_i)_{1 \leq i \leq n}$  can be chosen by statistician. We propose the special choice (depending on  $\mu$ ) of the design points that allows us to link (4.1) to WGN model. First we introduce required numerical tools for the univariate case. Then we apply this set-up to obtain optimal RNF in the derived d-dimensional model.

Actually. the method basically aims to select design point such as equivalent of (3.6) holds. So, let us decompose  $f$  with  $(\phi_k)_{k \in \mathbb{N}}$ , some arbitrary orthonormal basis of  $L_2(\mu)$  :

$$(4.2) \quad f(\cdot) = \sum_{k \in \mathbb{N}} \theta_k(f) \phi_k(\cdot).$$

Then still using projection method, we fit to  $f$  estimating part of sum of square of Fourier coefficients  $\theta_k(f) = \langle f, \phi_k \rangle$ . As usual, standard linear methods perform and for well chosen weighted positive coefficients  $(\lambda_i)$ ,

$$(4.3) \quad \sum_{i=1}^n \lambda_i f(x_i) \phi_k(x_i) \approx \theta_k(f).$$

We derive from such deterministic approximation the following family of weighted linear estimators

$$(4.4) \quad \hat{\theta}_k = \sum_{i=1}^n \lambda_i Y_i \phi_k(x_i) = \theta_k(f) + b_k(f) + \eta_k,$$

where

$$b_k(f) = \sum_{i=1}^n \lambda_i f(x_i) \phi_k(x_i) - \theta_k(f) \quad \text{and} \quad \eta_k = \sum_{i=1}^n \lambda_i \varepsilon_i \phi_k(x_i).$$

Our approach expects to give sufficient conditions on design points  $(x_i)_I$  as well as on projection basis  $(\phi_k)_k$  in order to related statistical treatment of (4.4) to these of (3.6).

**4.2. Choice of design points and projection basis. Univariate case.** The main difficulty arises in the suitable selection of design points  $(x_i)_{i=1}$ , weights  $(\lambda_i)_{i=1}$  and projection basis  $(\phi_k(\cdot))_{k \in \mathbb{N}}$ . For smooth function, standard numerical methods lead to estimate  $\theta_k(f)$  with the use of mean of  $f$ -values on a regular  $n$ -grid. For instance, if  $f$  is an odd and infinitely derivable function on  $[0, 1]$  and if  $(\phi_k)_k$  is the cosinus basis on  $[0, 1]$ , one would get a rather sharp approximation of  $\theta_k(f)$  using

$$\frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n}\right) \phi_k\left(\frac{i}{n}\right).$$

Though such choice provides good estimation of  $f$  under the assumption that  $\varepsilon_i \equiv 0$  in (4.1). Indeed, under noisy observations one may also control the stochastic part of estimation procedure. In a general regression setting, its study involves many technicalities. To avoid such disadvantages, we need to choose design points  $(x_i)$  and  $(\lambda_i)$  in order to keep under control in the same time the bias term  $b_k(f)$  (efficient integral approximation) and stochastic term  $\eta_k$ .

As remarked above, the main singularity of WGN model consists in (3.6) i.e. sequence of Fourier coefficients  $(\theta_k(f))_k$  can be observed with i.i.d. noise sample whatever the projection basis  $(\phi_k)_k$ . In the considered model, according to assumption on  $\varepsilon = (\varepsilon_i)$ , vector  $\eta = (\eta_k)_{k \in \mathbb{N}}$  is Gaussian. Thus, in order  $\eta$  to be an i.i.d. sample, it is sufficient to impose that  $\eta_k$  are not correlated, that is

$$(4.5) \quad \forall k \neq l, \quad E\eta_k \eta_l = \sum_{i=1}^n \lambda_i \phi_k(x_i) \phi_l(x_i) = 0.$$

According to numerical methods, one can linked condition (4.5) to the orthogonality of the basis  $(\phi_k)$ . Indeed for fixed  $n \in \mathbb{N}^*$  and a given positive function  $w(\cdot)$ , existing formulas, called quadrature formulas, allow to estimate integral of some function  $f$  by a weighted sum of a finite number, say  $n$ , values of  $f$  ( $n$ -points method).

$$(4.6) \quad \int_{\mathbb{R}} f(t) w(t) dt \simeq \sum_{i=1}^n \tau_i f(u_i),$$

where  $(u_i)$  and  $(\tau_i)$  do not depend on  $f$  but are linked with the function  $w(\cdot)$ . In this context, Gauss proved that there exists  $n$ -points method such that (4.6) is exact if  $f$  is polynomial function of degree lower than  $2n - 1$ . That is

$$(4.7) \quad \forall P \in \mathcal{P}_{2n-1}, \quad \int_{\mathbb{R}} P(t) w(t) dt = \sum_{i=1}^n \lambda_i P(x_i).$$

Namely let us define what we call in the following, the  $n^{th}$  orthogonal polynomial of  $L_2(\mu)$ . Using Gramm-Schmidt algorithm, we derive an orthonormal polynomial basis of  $L_2(\mu)$ , denoted by  $\Phi_w$ , from the canonical basis  $\{x^k; k \in \mathbb{N}\}$  of  $\mathcal{P}$ , set of univariate polynomials. Then, the  $n^{th}$  orthogonal polynomial of  $L_2(\mu)$  refers to the unique polynomial of  $\Phi_w$  that have degree  $n$ .

Hence, (4.7) is satisfied for  $x_w^{(n)} = (x_i; i = 1, \dots, n)$  as the zeros of the  $n^{th}$  orthogonal polynomial of  $L_2(\mu)$ . This choice is optimal in the sense that there exists no  $n$ -points method that is exact on  $\mathcal{P}_{2n}$ . Note that it can be proved that  $x_w^{(n)} \subset \mathbb{R}$  and then  $x_i$  are real numbers. In what follows, the grid  $x_w^{(n)}$  that fundamentally depends on the choice  $w(\cdot)$ , is referred to the Gauss grid.

This general result will lead us to project the underlying function  $f$  on the *orthonormal polynomial basis* of  $L_2(\mu)$ ,  $(\phi_k)_{k \in \mathbb{N}} = \Phi_w$ . Thus for  $1 \leq k \neq l \leq n - 1$ , condition (4.5) is fulfilled since

$$(4.8) \quad \sum_{i=1}^n \lambda_i \phi_k(x_i) \phi_l(x_i) = \int_{\mathbb{R}} \phi_k(t) \phi_l(t) w(t) dt = \langle \phi_k, \phi_l \rangle = 0.$$

**Remark 4.1.** Note that the construction of this specific observation grid  $x_w^{(n)}$  as well as the choice of projection basis  $\Phi_w$  heavily depends on function  $w(\cdot)$ .

**Remark 4.2.** Below, we will see that projection estimation procedure only requires to estimate  $\theta_k(f)$  for  $k \leq N_n$ , where typically  $N_n \leq n^c$  with  $0 < c < 1$ . Thus, design points  $(x_i)$  can be chosen such that at least (4.8) holds for  $1 \leq k \neq l \leq n^c$ . However, for other reasons we will observe  $f$  using the optimal Gauss grid.

**Remark 4.3.** This procedure can be viewed as follows. First we project the regression function  $f$  over the polynomial functions of  $L_2(\mu)$  using Lagrange interpolation operator on the observation grid, denoted by  $\mathcal{J}_{n-1}f$ . According to what precedes, we observe Fourier coefficient of  $\mathcal{J}_{n-1}f$  in the equivalent of the WGN model (3.6). In a last step, we use orthogonal projection of  $\mathcal{J}_{n-1}f$ , on the subspace of additive function. Such procedure leads to an additional bias term that arises from the interpolation step. The choice of optimal Gauss design points would provide in some particular case relevant control of this bias term (linked with  $b_k(f)$ ) w.r.t. to the smoothness of function  $f$ .

**4.3. Multidimensional set-up.** Let  $f$  be a  $d$ -dimensional function ( $d > 1$ ) from  $L_2(\underline{\mu})$  when  $\underline{\mu} = \bigotimes_{i=1}^d \mu_i$  is a product measure on  $\mathbb{R}^d$  and  $\mu_i$  are absolutely continuous w.r.t. to Lebesgue measure  $\lambda$  with  $w_i(\cdot) = \frac{\partial \mu_i}{\partial \lambda}(\cdot)$ . In a care of simplicity, we would assume that  $\mu_i = \mu$ , then  $w_i(\cdot) = w(\cdot)$ . This technical restriction does not prevent to extend the procedure to a more general framework. Thus, we put

$$\mathbf{w}(t) = \prod_{i=1}^d w(t_i), \quad \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Denote by  $\|\cdot\|_{L_2(\underline{\mu})}$  the Hilbert norm on  $L_2(\underline{\mu})$ . Assume that  $f$  is observed in the  $d$ -dimensional regression model with selected design derived from (4.1). One denotes by  $\Phi_w^{(d)} = (\phi_k^{(d)}, k \in \mathbb{N}^d)$  the total family of  $L_2(\underline{\mu})$  where

$$(4.9) \quad \forall k = (k_1, \dots, k_d) \in \mathbb{N}^d, \quad \phi_k^{(d)}(t) = \prod_{i=1}^d \phi_{k_i}(t_i),$$

when  $(\phi_k)_{k \in \mathbb{N}} = \Psi_w$  is defined above.

For any  $k \in \mathbb{N}^d$ , set  $\theta_k(f) = \langle \phi_k^{(d)}, f \rangle = \int_{\mathbb{R}^d} f(t) \phi_k^{(d)}(t) \mathbf{w}(t) dt$ . Then for any  $f \in L_2(\underline{\mu})$ , the following expansion holds w.r.t.  $\|\cdot\|_{L_2(\underline{\mu})}$ ,

$$f(\cdot) = \sum_{k \in \mathbb{N}^d} \theta_k(f) \phi_k^{(d)}(\cdot).$$

For a fixed  $m \in \mathbb{N}^*$ , and  $L > 0$  we suppose that  $f$  belongs to the Sobolev ellipsoid

$$(4.10) \quad \Theta_w = \Theta_w(m, L) = \left\{ f \in L_2(\underline{\mu}) : \sum_{k \in \mathbb{N}^d} \theta_k^2 \left( 1 + \sum_{i=1}^d k_i^m \right) \leq L \right\}.$$

**Remark 4.4.** As in WGN model, considering such class of functions will permit to keep under control the bias deviation that corresponds to the projection estimation. This condition ensures that estimating  $f$  by its Fourier expansion performs well.

For a given  $M \in \mathbb{N}^*$ , we define the  $d$ -dimensional Gauss grid of size  $M^d$ ,

$$(4.11) \quad \Xi_{w,M} = \left\{ \mathbf{x}_i = (x_{i_1}, \dots, x_{i_d}) \in \mathbb{R}^d : \forall j \in \{1, \dots, d\}, x_{i_j} \in x_w^{(M)} \right\}.$$

Put  $\{L_i; i \in \{1, \dots, M\}^d\}$  the set of  $d$ -dimensional Lagrange polynomials attached to the grid  $\Xi_M$  i.e.  $L_i$  are  $d$ -dimensional polynomials such that

$$\forall \mathbf{i} \neq \mathbf{j} \in \{1, \dots, M\}^d, \quad L_i(x_j) = \delta_{\mathbf{i}, \mathbf{j}},$$

where  $\delta_{\mathbf{i}, \mathbf{j}}$  refers to the Kronecker operator on  $\mathbb{N}^d$ . Thus, let us define

$$\forall \mathbf{i} \in \{1, \dots, M\}^d, \quad \lambda_i = \int_{\mathbb{R}^d} L_i^2(t) \mathbf{w}(t) dt.$$

Hence, d-dimensional version of Gauss quadrature formula can be expressed as follows

$$(4.12) \quad \forall P \in \mathcal{P}_{2M-1}(d), \quad \sum_{x_i \in \Xi_{w,M}} \lambda_i P(x_i) = \int_{\mathbb{R}^d} P(t) \mathbf{w}(t) dt,$$

when  $\mathcal{P}_N(d)$  denotes the space of d-dimensional polynomials with degree lower than  $N$  i.e. functions  $P$  that can be written as

$$(4.13) \quad P(x) = \sum_{k \in \mathbb{N}^d} a_k x_1^{k_1} \dots x_d^{k_d},$$

with  $N \geq \max \{\|k\|_\infty : a_k \neq 0\}$  when  $\|k\|_\infty = \max_{i=1,\dots,d} (k_i)$ .

**Remark 4.5.** Note that it may happen for some  $w(\cdot)$  that exact values of design points  $(x_i; i = 1 \dots M)$  would not be directly available. Nevertheless, once  $M$  is fixed,  $\phi_M$  is provided by Gramm-Schmidt algorithm, and then some simple numerical (Newton-Raphson or Laguerre) methods enables to fit to  $(x_i)$  and then to  $(\lambda_i)$  with an arbitrary accuracy. According to the high smoothness of basis functions (polynomials), the use of design arbitrary closed to Gauss points will not damage the performance of the method. We will see further that contrary to the Chebychev grid, nodes of Legendre grid on  $[-1, 1]^d$  can not be exactly computed.

**Condition 2.** Assume that there exists  $c_1 > 0$  such that

$$\forall M \in \mathbb{N}^*, \quad \forall \mathbf{i} \in \{1, \dots, M\}^d, \quad 0 < \lambda_i \leq c_1 M^{-d}.$$

**Remark 4.6.** This condition involves that the weight coefficients are rather "regular". Indeed it states that optimal quadrature formula is obtained for non degenerate weight sequence  $\lambda_i$ .

Let denote by  $\mathcal{J}_{M-1}$  the Lagrange interpolation operator on  $\Xi_{w,M}$  defined by

$$\forall v \in L_2(\underline{\mu}), \quad \mathcal{J}_{M-1}(v)(\cdot) = \sum_{\mathbf{i} \in \{1, \dots, M\}^d} v(x_i) L_i(\cdot)$$

Remark that for any function  $v \in L_2(\underline{\mu})$ ,  $\mathcal{J}_{M-1}v$  belongs to  $\mathcal{P}_{M-1}(d)$ . We expect that function  $f$  can be approached by its interpolator  $\mathcal{J}_{M-1}(f)$  with sufficient accuracy. To measure the quality of interpolation w.r.t. the number of observation points, we define for any  $m \in \mathbb{N}^*$  and any  $c > 0$ ,

$$(4.14) \quad \Upsilon_w(m, c) = \left\{ v \in L_2(\underline{\mu}) : \forall M \in \mathbb{N}^*, \quad \|v - \mathcal{J}_{M-1}v\|_{L_2(\underline{\mu})} \leq cM^{-m} \right\}.$$

**Remark 4.7.** We are allowed to control the interpolation error uniformly over  $\Upsilon_w(m, c)$ . Such class of functions is essential in our procedure. Indeed considering  $f \in \Upsilon_w(m, c)$  would permit to state that observing  $(\theta_k(\mathcal{J}_{M-1}f))$  instead of  $(\theta_k(f))$  does not reduce efficiency of estimation procedure. Actually, this condition would guarantee that 'interpolation' bias attached to sequence  $(b_k(f))_k$  can be neglected. Consequently, it would still be relevant to test additive structure for  $f$  estimating  $\sum_{k \in \mathbb{N}^d: |G_k| > 1} \theta_k^2(\mathcal{J}_{M-1}f)$  where  $G_k$  is defined as in section 3.

**4.4. Notations.** Let  $M_n = n^{\frac{1}{d}}$ . Without loss of generality we will assume that  $M_n$  is an integer. We choose to observe the regression function on the grid  $\Xi_{M_n}$ . Remark that this choice involves exactly  $n$  design points. Then one observes

$$(4.15) \quad Y_i = f(x_i) + \varepsilon_i, \quad \forall x_i \in \Xi_{M_n}.$$

when  $\varepsilon_i$  are i.i.d Gaussian zero-mean noise variables with variance  $\sigma > 0$ . Assume that there exists  $m \in \mathbb{N}^*$ ,  $c > 0$  and  $L > 0$  such that

$$(4.16) \quad f \in \Sigma = \Sigma_w(m, L, c) = \Theta_w(m, L) \cap \Upsilon_w(m, c).$$

The subset of functions satisfying to structure (1.5) will be denoted by  $\Sigma_0 = \Sigma_{0,w}(m, L, c)$ .

**Remark 4.8.** Introduction of  $\Sigma_w(m, L, c)$  enables to control in the same time the projection estimation error inherent to stastical nonparametric projection method, but also the quality of interpolation. Both allows to get a reasonable bias deviation in the estimation of  $\|f\|_{L_2(\underline{\mu})}$ .

**Remark 4.9.** Below, we treat two particular cases for function  $w(\cdot)$ , for which one observes that considering  $\Sigma$  defined as in 4.16 is hardly restrictive.

Mathematical description of estimation approach with RNF is derived from WGN model to regression model. By analogy, we measure quality of estimation replacing the risk (1.2) by

$$R_n^{(r)}(\hat{f}_n, \Sigma, \rho_n) = \sup_{f \in \Sigma} E_f \{ \rho_n^{-1} \|f - \hat{f}_n\|_{L_2(\mu)} \}^2,$$

where  $\rho_n = \{b_n, \varphi_n(\Sigma)\}$  is a RNF from  $\mathfrak{E}_n$  set of bounded RNF that takes only two values  $\{\varphi_n(\Sigma), b_n\}$  with  $0 < b_n < \varphi_n(\Sigma)$ .

For any  $k \in \mathbb{N}^d$ , if  $\{\lambda_{i,n} : i \in \{1, \dots, M_n\}^d\}$  are Gauss weights attached to the grid  $\Xi_{M_n}$ , set

$$(4.17) \quad \hat{\theta}_k = \sum_{x_i \in \Xi_n} \lambda_{i,n} Y_i \phi_k^{(d)}(x_i).$$

Then the following decomposition holds :

$$(4.18) \quad \hat{\theta}_k = \theta_k(\mathcal{J}_{n-1}f) + \eta_k = \theta_k(f) + b_k(f) + \eta_k,$$

where  $\mathcal{J}_{n-1} = \mathcal{J}_{M_n-1}$  is the Lagrange interpolation operator on  $\Xi_{M_n}$  and

$$\begin{aligned} \eta_k &= \sum_{x_i \in \Xi_n} \lambda_{i,n} \varepsilon_i \Phi_k(x_i), \\ b_k(f) &= \theta_k(\mathcal{J}_{n-1}f) - \theta_k(f) = \sum_{x_i \in \Xi_n} \lambda_{i,n} f(x_i) \Phi_k(x_i) - \int_{[-1,1]^d} f(t) \Phi_k(t) dt. \end{aligned}$$

Here  $b_k(f)$  corresponds to the additional bias term induced by the interpolation step. It measures error made confusing  $\theta_k(f)$  and  $\theta_k(\mathcal{J}_{n-1}f)$ . Note that variable  $\eta_k$  is a Gaussian random variable with zero mean and variance  $\sigma_k^2 = \sum_i \lambda_{i,n}^2 \Phi_k^2(x_i)$ .

**4.5. Construction and results.** Now let us fixe  $0 < \alpha_n < 1$ , then we define

$$\begin{aligned} \Gamma_n &= \{k \in \mathbb{N}^d : k_j < N_n \ \forall j = 1\} \quad \text{where} \quad N_n = n^{1/(2m+d)}, \\ \Delta_n &= \{k \in \mathbb{N}^d : k_j < N_{0,n} \ \forall j = 1\} \quad \text{where} \quad N_{0,n} = 2 \left( n^{-1} \sqrt{\ln(1/\alpha_n)} \right)^{-2/(4m+d)}. \end{aligned}$$

If for any  $k \in \mathbb{N}^d$ ,  $G_k$  is defined as in (3.9), then we set

$$\begin{aligned} I_n &= \Delta_n \cap \{k \in \mathbb{N}^d : |G_k| \leq 1\}, \\ \Lambda_n &= \Delta_n \setminus I_n. \end{aligned}$$

We consider the following estimator :  $\hat{f}_n = \sum \hat{\theta}_{n,k} \Phi_k$  and  $\hat{f}_n^0 = \sum \hat{\theta}_{n,k}^0 \Phi_k$  when

$$\hat{\theta}_{n,k} = \begin{cases} \hat{\theta}_k, & \text{if } k \in \Gamma_n, \\ 0, & \text{if not.} \end{cases} \quad \hat{\theta}_{\varepsilon,k}^0 = \begin{cases} \hat{\theta}_k, & \text{if } k \in I_n, \\ 0, & \text{if not.} \end{cases}$$

The decision rule of the test corresponds to the occurrence or not of the event

$$A_n = \{T_n \leq \lambda a_n^2\},$$

when  $\lambda$  only depends on  $(m, d)$ , and

$$a_n = \left( n^{-1} \sqrt{\ln(1/\alpha_n)} \right)^{2m/(4m+d)} \quad \text{and} \quad T_n = \sum_{k \in \Lambda_n} \left( \hat{\theta}_k^2 - \sigma_k^2 \right).$$

Finally, we define the RNF  $\rho_n^*$  and the estimator  $f_n^*(\cdot) = \sum_{k \in \mathbb{N}^d} \theta_{n,k}^* \Phi_k(\cdot)$  by

$$\rho_n^* = \begin{cases} a_n, & \text{on } A_n, \\ \varphi_n = n^{-m/(2m+d)}, & \text{if not,} \end{cases} \quad \text{and} \quad f_n^* = \begin{cases} \hat{f}_n^0, & \text{on } A_n, \\ \hat{f}_n, & \text{if not,} \end{cases}$$

By analogy with (2.2),

**Proposition 4.1.** *The RNF  $(\rho_n^*)_n$  satisfies to*

$$(4.19) \quad \limsup_{n \rightarrow +\infty} \alpha_n^{-1} \sup_{f \in \Sigma_0} P_f^n \{ \rho_n^* = \varphi_n \} \leq 1$$

Note that under condition that  $\varphi_n$  is the MRC on  $\Sigma$ , then proposition 4.1 is analogue to lemma 3.1.

**Condition 3.** *The sequence  $(\alpha_n)_n$  satisfies to*

$$\exists a > 0 \quad : \quad \forall n \geq 1, \quad 1 > \alpha_n > n^{-a}.$$

**Theorem 4.1.** *Under conditions 2 and 3 and if (4.16) holds with  $m > \frac{d}{4}$ , the estimator  $f_n^*$  is adapted to the RNF  $\rho_n^*$  i.e. there exist  $C = C(c, m, d, L) > 0$  such that*

$$\lim_{n \rightarrow \infty} R_n^{(r)}(f_n^*, \Sigma, \rho_n^*) < C < \infty.$$

We leave to the reader the complete proof of this theorem. In section 5, we just linked computations in this regression model with what have been already proved in W.G.N. model.

**Remark 4.10.** *One conjectures that under assumption of theorem 4.1 and if  $\varphi_n(\Sigma) = \varphi_n$ , then  $(\rho_n^*)_n$  may be  $\alpha$ -optimal w.r.t.  $\{\Sigma, \Sigma_0\}$ . Indeed proposition 4.1 proves that  $\rho_n^* \in \Omega(\alpha)$  and theorem 4.1 gives the upper bound, first condition of RNF optimality definition 2.2. We presume that second condition is satisfied for this specified regression model.*

One may wonder if condition  $f \in \Sigma_w(m, L, c)$  is not too restrictive. Indeed one may insure that such kind of set are not too small. One can not answer this question under general consideration. Though there exist results [2] that prove that for particular function  $w(\cdot)$ , consideration of these class of function is reasonable. Last parts of this section is devoted to the presentation of two specific frameworks. In both case, one can measure interpolation error in accordance with some smoothness parameter. Indeed, with same notations as above put

$$W_w(m, c) = \left\{ f \in L_2(\underline{\mu}) : \|f\|_{W_w^m}^2 = \sum_{|\beta| \leq m} \int_{[-1,1]^d} (\partial^\beta f)^2(t) \mathbf{w}(t) dt < c \right\}.$$

This class of function is rather closed to ellipsoid  $\Theta_w(m, L)$ . In particular, underline that under periodicity assumption they may be equal. We present two choice of  $w(\cdot)$  where  $W_w(m, c) \subset \Upsilon_w(m, c)$ .

**4.6. Study of Legendre design.** Assume that  $\mathbf{w}(t) = \mathbf{w}_1(t) = \mathbf{1}_{[-1,1]^d}(t)$ , that is  $w(t) = w_1(t) = \mathbf{1}_{[-1,1]}(t)$ . Thus, the risk (4.17) corresponds to the usual quadratic risk on  $[-1, 1]^d$ . Orthonormal d-dimensional polynomial basis is given by

$$\forall k \in \mathbb{N}^d, \quad \phi_k^{(d)}(t) = \mathbf{L}_k(t) = \prod_{i=1}^d L_{k_i}(t_i),$$

where  $L_k$  are unidimensional Legendre polynomials. Denote by  $\{\xi_j = \cos(\vartheta_j); j = 1 \dots M\}$  the M distincts zeros of  $L_M$ . Then there exists M real numbers  $\omega_j$ ,  $1 \leq j \leq M$  such that

$$\forall P \in \mathcal{P}_{2M-1}(d), \quad \int_{[-1,1]^d} P(t) dt = \sum_{\mathbf{i} \in \{1, \dots, M\}^d} \left( \prod_{s=1}^d \omega_{i_s} \right) P(\xi_{i_1}, \dots, \xi_{i_d}).$$

Moreover, Szego ([31]) gives a good estimation for  $\vartheta_j$  and  $\omega_j$ , since he proves that there exists  $c, c' > 0$  such that

$$(4.20) \quad \forall 1 \leq j \leq M, \quad \frac{(M - j + \frac{1}{2})\pi}{M} \leq \vartheta_j \leq \frac{(M - j + 1)\pi}{M},$$

$$(4.21) \quad cM^{-1} \sqrt{1 - \xi_j^2} \leq \omega_j \leq c'M^{-1} \sqrt{1 - \xi_j^2}.$$

Let us remark that (4.21) involves condition 2.

**Remark 4.11.** *As underlined above, in this particular case exact values of  $\xi_j$  and  $\omega_j$  are not available. Nevertheless, numerical algorithms such as Newton, Laguerre or Givens-Householder (see [5]) enable to give with more or less efficiency, approximation of their value with an arbitrary accuracy i.e. that doesn't depend on  $M$  [2].*

Using the optimality of Gauss quadrature formula as well as (4.20), Bernardi and Maday [2] proved that there exist an constant  $c^* > 0$  that only depends on  $w_1(\cdot)$  such that

$$(4.22) \quad \forall v \in L_2(\underline{\mu}), \quad \forall M \in \mathbb{N}, \quad \|v - \mathcal{J}_{M-1}(v)\|_{L_2} \leq c^* M^{-m} \|v\|_{W_w^m}^2$$

when  $\mathcal{J}_{M-1}(v)$  is the interpolation operator on the d-grid whose node are  $\{\xi = (\xi_{i_s})_{s=1}^d : 1 \leq i_s \leq M\}$ . One can deduce from this result that for any  $c > 0$ ,  $W_{w_1}(m, c) \subset \Upsilon_{w_1}(m, cc^*)$ . Then one can apply theorem 4.1 when  $\Sigma = W_{w_1}(m, c) \cap \Theta_{w_1}(m, L)$ .

**4.7. Study of Chebychev design.** Now let  $\mathbf{w}(t) = \mathbf{w}_2(t) = \mathbf{1}_{[-1,1]^d}(t) \prod_{i=1}^d \frac{1}{\sqrt{1-t_i^2}}$  be the weight function. In this context,

$$\forall k \in \mathbb{N}^d, \quad \phi_k^{(d)}(t) = \mathbf{T}_k(t) = \prod_{i=1}^d T_{k_i}(t_i),$$

where  $T_k(\xi) = \cos(k \arccos(\xi))$  are unidimensional Tchebychev polynomials. In this particular case, exact expression of zeros of  $T_M$  are available :

$$\forall j = 1 \dots M, \quad \xi_j^u = \cos\left(\frac{(M-j+\frac{1}{2})\pi}{M}\right).$$

Denote by  $\mathcal{J}_{M-1}^u$  the operator on the d-dimensional grid  $\{\xi = (\xi_{i_s}^u)_{s=1}^d : 1 \leq i_s \leq M\}$ . As well as  $\xi_j^u$ , weight coefficients can be exactly computed :  $\forall j = 1, \quad \omega_j^u = \frac{\pi}{N}$ . Thus condition 2 is fulfilled.

Similarly, it is proved in [2] that there exists  $c^{**} = c^{**}(w_2) > 0$  such that for any  $c > 0$ ,  $W_{w_2}(m, c) \subset \Upsilon_{w_2}(m, cc^{**})$ . Then equivalent results can be stated with  $\Sigma = W_{w_2}(m, c) \cap \Theta_{w_2}(m, L)$ .

**Remark 4.12.** *One can extend these results to any measure  $\underline{\mu}$  with compact support in  $\mathbb{R}^d$ . For example, if  $\mathbf{w}(t) = \mathbf{1}_{[-a,a]^d}(t)$  or  $\mathbf{w}(t) = \mathbf{1}_{[-a,a]^d}(t) \prod_{i=1}^d \frac{1}{\sqrt{a-t_i^2}}$ , changing  $t$  to  $t' = t/a$  enables to bring back to the considered models.*

## 5. PROOFS

**5.1. Proof of lemma 3.1.** It is sufficient to check that

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{\theta \in \Sigma_0} P_\theta(\rho_\varepsilon^* \neq a_\varepsilon) \leq 1.$$

So if  $\theta \in \Sigma_0$ , we have :

$$P_\theta(\rho_\varepsilon^* \neq a_\varepsilon) = P_\theta(A_\varepsilon^c) = P_\theta(T_\varepsilon > \lambda a_\varepsilon^2),$$

when according to (3.10),

$$T_\varepsilon = \varepsilon^2 \sum_{k \in \Lambda_\varepsilon} (\zeta_k^2 - 1)$$

is the centered sum of square of i.i.d. standard gaussian random variables. In this case, large deviations are well known since

$$(5.1) \quad \begin{aligned} \forall t > 0, \quad P_\theta(T_\varepsilon > \lambda a_\varepsilon^2) &\leq \exp^{-t\lambda a_\varepsilon^2} E \left[ e^{t\varepsilon^2(\zeta^2-1)} \right]^{|\Lambda_\varepsilon|} \\ &\leq \exp^{-t\lambda a_\varepsilon^2} \left[ \sum_{p=0}^{\infty} \frac{t^p \varepsilon^{2p}}{p!} E(\zeta^2 - 1)^p \right]^{|\Lambda_\varepsilon|}. \end{aligned}$$

Note  $\mu = \frac{1}{2} E(\zeta^2 - 1)^2$  and let  $c > 0$  an nonnegative real such that

$$\forall p \geq 3, \quad E(\zeta^2 - 1)^p \leq c^p p!.$$

Then if  $\gamma$  is fixed real such as  $0 < \gamma < 1$ , for any  $t \leq \gamma c^{-1} \varepsilon^{-2}$ , (5.1) becomes

$$\begin{aligned} P_\theta(T_\varepsilon > \lambda a_\varepsilon^2) &\leq e^{-t\lambda a_\varepsilon^2} \left[ 1 + \mu t^2 \varepsilon^4 \left\{ 1 + \sum_{p=3}^{\infty} \frac{(t\varepsilon^2)^{p-2}}{\mu p!} E(\zeta^2 - 1)^p \right\} \right]^{|\Lambda_\varepsilon|} \\ &\leq e^{-t\lambda a_\varepsilon^2} \left[ 1 + \mu t^2 \varepsilon^4 \left\{ 1 + \frac{c^2}{\mu} \sum_{p=3}^{\infty} (ct\varepsilon^2)^{p-2} \right\} \right]^{|\Lambda_\varepsilon|} \\ &\leq e^{-t\lambda a_\varepsilon^2} \left[ 1 + \mu t^2 \varepsilon^4 \left\{ 1 + \frac{c^2 C_\gamma}{\mu} \right\} \right]^{|\Lambda_\varepsilon|}. \end{aligned}$$

Using that  $\forall y \in \mathbb{R}, 1 + y \leq e^y$ , and from  $|\Lambda_\varepsilon| \leq N_{0,\varepsilon}^d$ , it follows that

$$P_\theta(T_\varepsilon > \lambda a_\varepsilon^2) \leq \exp \left\{ -t\lambda a_\varepsilon^2 + \mu \left( 1 + \frac{c^2 C_\gamma}{\mu} \right) t^2 \varepsilon^4 N_{0,\varepsilon}^d \right\}.$$

Minimum of the right hand side term w.r.t.  $t > 0$ , is attained for  $t_\varepsilon \asymp \frac{\varepsilon^{-4} a_\varepsilon^2}{N_{0,\varepsilon}^d} \ll \varepsilon^{-2}$ , so when we set  $Q = Q(\mu, \gamma) = 2^{d+2} \mu \left( 1 + \frac{c^2 C_\gamma}{\mu} \right)$  and  $\lambda \geq \sqrt{Q}$  the previous inequality becomes :

$$P_\theta(\rho_\varepsilon^* \neq a_\varepsilon) \leq \exp \left( -\frac{\lambda^2}{Q} \ln(1/\alpha_\varepsilon) \right) = \alpha_\varepsilon^{\frac{\lambda^2}{Q}} \leq \alpha_\varepsilon.$$

## 5.2. Proof of theorem 3.1.

5.2.1. *Upper bound.* For any  $\theta \in \Sigma$ , let us decompose the error in the estimation of a function  $f \in \Sigma$  according to the outcome of the test. Set :

$$R_\varepsilon^{(1)}(\theta) = E_\theta \left\{ a_\varepsilon^{-2} \|\widehat{\theta}_\varepsilon^0 - \theta\|^2 1_{A_\varepsilon} \right\},$$

$$R_\varepsilon^{(2)}(\theta) = E_\theta \left\{ \varphi_\varepsilon(\Sigma)^{-2} \|\widehat{\theta}_\varepsilon - \theta\|^2 1_{A_\varepsilon^c} \right\}.$$

Minimax risk with RNF on  $\Sigma$  is given by

$$\begin{aligned} R_\varepsilon^{(r)}(\theta_\varepsilon^*, \rho_\varepsilon^*) &= \sup_{\theta \in \Sigma} E_\theta \left\{ (\rho_\varepsilon^*)^{-2} \|\theta^* - \theta\|^2 \right\} \\ (5.2) \quad &\leq \sup_{\theta \in \Sigma} R_\varepsilon^{(1)}(\theta) + \sup_{\theta \in \Sigma} R_\varepsilon^{(2)}(\theta). \end{aligned}$$

Obviously, choice of  $\widehat{\theta}_\varepsilon$  as a minimax estimator on  $\Sigma$  allows to control the second term as :

$$\begin{aligned} R_\varepsilon^{(2)}(\theta) &\leq \varphi_\varepsilon(\Sigma)^{-2} E_\theta \left\{ \|\widehat{\theta}_\varepsilon - \theta\|^2 \right\} \\ &\leq \varphi_\varepsilon(\Sigma)^{-2} E_\theta \left\{ \varepsilon^2 \sum_{k \in \Gamma_\varepsilon} \zeta_k^2 + \sum_{k \notin \Gamma_\varepsilon} \theta_k^2 \right\} \\ &\leq \varphi_\varepsilon(\Sigma)^{-2} \left( \varepsilon^2 N_\varepsilon^d + \sum_{k \notin \Gamma_\varepsilon} \theta_k^2 \right) \leq \varphi_\varepsilon(\Sigma)^{-2} \left( \varepsilon^2 N_\varepsilon^d + \frac{L}{N_\varepsilon^{2\beta}} \right). \end{aligned}$$

Then the choice of parameters of our procedure induces that

$$(5.3) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{\theta \in \Sigma} R_\varepsilon^{(2)}(\theta) \leq 1 + L.$$



Now we will interest in the term connected with the acceptance of the hypothesis of additivity.

$$\begin{aligned} R_\varepsilon^{(1)}(\theta) &= a_\varepsilon^{-2} E_\theta \left\{ \left( \varepsilon^2 \sum_{k \in I_\varepsilon} \zeta_k^2 + \sum_{k \notin I_\varepsilon} \theta_k^2 \right) 1_{A_\varepsilon} \right\} \\ &= a_\varepsilon^{-2} E_\theta \left\{ \left( \varepsilon^2 \sum_{k \in I_\varepsilon} \zeta_k^2 + H_\varepsilon(\theta) + K_\varepsilon(\theta) \right) 1_{A_\varepsilon} \right\} \end{aligned}$$

where

$$K_\varepsilon(\theta) = \sum_{k \notin \Delta_\varepsilon} \theta_k^2 \quad \text{and} \quad H_\varepsilon(\theta) = \sum_{k \in \Lambda_\varepsilon} \theta_k^2.$$

Since  $\theta$  belongs to  $\Sigma$ , we get the upper bound for the bias  $K_\varepsilon(\theta) \leq L^2 a_\varepsilon^2$ . Moreover as  $|I_\varepsilon| = d(N_{0,\varepsilon} - 1)$ , that yields

$$(5.4) \quad R_\varepsilon^{(1)}(\theta) \leq a_\varepsilon^{-2} \{ L^2 a_\varepsilon^2 + d\varepsilon^2 N_{0,\varepsilon} + H_\varepsilon(\theta) \} P_\theta(A_\varepsilon).$$

The event  $A_\varepsilon$  depends on the statistic  $T_\varepsilon$  that could be expended in the following way

$$T_\varepsilon = H_\varepsilon(\theta) + \eta_\varepsilon(\theta) + \varepsilon^2 S(\Lambda_\varepsilon),$$

when

$$\eta_\varepsilon(\theta) = 2\varepsilon \sum_{k \in \Lambda_\varepsilon} \theta_k \zeta_k$$

is a random variable with a centered gaussian law  $\mathcal{N}(0, 4\varepsilon^2 H_\varepsilon(\theta))$ ; and when we defined for  $\mathcal{A} \subset \mathbb{N}^d$ , the sum

$$S(\mathcal{A}) = \sum_{k \in \mathcal{A}} (\zeta_k^2 - 1).$$

To bound the risk (5.4) uniformly over  $\Sigma$ , we will decompose  $\Sigma$  w.r.t. value of deviation  $H_\varepsilon(\theta)$ . Namely for  $\delta > 0$ , set  $\Theta_{\varepsilon,\delta}$  the subspace of  $\Sigma$  defined by :

$$\Theta_{\varepsilon,\delta} = \left\{ \theta \in \Sigma : H_\varepsilon(\theta) \leq \frac{1+\delta}{1-\delta} \lambda^2 a_\varepsilon^2 \right\}.$$

Thus let decompose this risk as follows :

$$(5.5) \quad \sup_{\theta \in \Sigma} R_\varepsilon^{(1)}(\theta) \leq R_\varepsilon^{(1,1)} + R_\varepsilon^{(1,2)},$$

when

$$R_\varepsilon^{(1,1)} = \sup_{\theta \in \Theta_{\varepsilon,\delta}} R_\varepsilon^{(1)}(\theta), \quad \text{and} \quad R_\varepsilon^{(1,2)} = \sup_{\theta \in \Sigma - \Theta_{\varepsilon,\delta}} R_\varepsilon^{(1)}(\theta).$$

Using (5.4) and since  $a_\varepsilon^{-2} \varepsilon^2 N_{0,\varepsilon} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , we get

$$(5.6) \quad \limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(1,1)} \leq L^2 + \frac{1+\delta}{1-\delta} \lambda^2.$$

Concerning the second term  $R_\varepsilon^{(1,2)}$ , we need to remark that for  $\theta \in \Sigma - \Theta_{\varepsilon,\delta}$ , Chebychev exponential inequality ensures that

$$P_\theta(|\eta_\varepsilon(\theta)| \geq \delta H_\varepsilon(\theta)) \leq 2 \exp \left\{ -\frac{\delta^2 \varepsilon^{-2} H_\varepsilon(\theta)}{2} \right\}.$$

Then there exists a constant  $\gamma > 0$  independent from  $\lambda$  and  $\theta \in \Sigma$  such that

$$P_\theta\{|\eta_\varepsilon(\theta)| \geq \delta H_\varepsilon(\theta)\} \leq \exp\{-\varepsilon^{-2\gamma}\}.$$

It follows that  $\forall \theta \in \Sigma - \Theta_{\varepsilon,\delta}$ ,

$$(5.7) \quad P_\theta(A_\varepsilon) = P_\theta(T_\varepsilon \leq \lambda^2 a_\varepsilon^2) \leq \exp\{-\varepsilon^{-2\gamma}\}(1 + o(1)) + P_\theta(H_\varepsilon(1-\delta) + \varepsilon^2 S(\Lambda_\varepsilon) \leq \lambda^2 a_\varepsilon^2).$$

Both (5.4) and (5.7) yield

$$(5.8) \quad R_\varepsilon^{(1,2)} \leq L^2 + \sup_{\theta \in \Sigma - \Theta_{\varepsilon,\delta}} F_\varepsilon(\theta) + o(1),$$

when for any  $\theta \in \Sigma$  we define

$$F_\varepsilon(\theta) = a_\varepsilon^{-2} H_\varepsilon(\theta) P_\theta (H_\varepsilon(\theta)(1 - \delta) + \varepsilon^2 S(\Lambda_\varepsilon) \leq \lambda^2 a_\varepsilon^2).$$

Now let us decompose  $\Sigma - \Theta_{\varepsilon,\delta} = \bigcup_{x \geq 1+\delta} \Theta(x)$  where we define

$$\Theta(x) = \left\{ \theta \in \Sigma : \frac{H_\varepsilon(\theta)(1 - \delta)}{\lambda^2 a_\varepsilon^2} = x \right\}.$$

It leads to consider  $\vartheta_\varepsilon : x \longrightarrow \sup_{\theta \in \Theta(x)} F_\varepsilon(\theta)$ . Still using exponential Chebychev inequality one have

$$\begin{aligned} \vartheta_\varepsilon(x) = \sup_{\theta \in \Theta(x)} F_\varepsilon(\theta) &\leq \frac{\lambda^2 x}{1 - \delta} P_\theta (S(\Lambda_\varepsilon) \leq -(x - 1) \lambda^2 a_\varepsilon^2 \varepsilon^{-2}) \\ &\leq \frac{\lambda^2 x}{1 - \delta} \exp \left( - \frac{\varepsilon^{-4} (x - 1)^2 \lambda^4 a_\varepsilon^4}{\Lambda_\varepsilon} \right) \\ &\leq \frac{\lambda^2 x}{1 - \delta} \exp \left( -(x - 1)^2 \ln \left( \frac{1}{\alpha_\varepsilon} \right) \right). \end{aligned}$$

As  $\alpha_\varepsilon \rightarrow 0$  when  $\varepsilon \rightarrow 0$ , suppose that  $\varepsilon$  is small enough and  $\alpha_\varepsilon < \frac{1}{2}$ . That provides the upper bound,

$$\begin{aligned} \sup_{\theta \in \Sigma - \Theta_{\varepsilon,\delta}} F_\varepsilon(\theta) &= \sup_{x \geq 1+\delta} \vartheta_\varepsilon(x) \leq \sup_{x \geq 1+\delta} \frac{\lambda^2 x}{1 - \delta} \exp \left( -(x - 1)^2 \ln \left( \frac{1}{\alpha_\varepsilon} \right) \right) \\ &\leq \sup_{x \geq \delta} \frac{\lambda^2 (x + 1)}{1 - \delta} \exp \left( -x^2 \ln \left( \frac{1}{\alpha_\varepsilon} \right) \right) \\ (5.9) \quad &\leq \frac{\lambda^2}{2\delta(1 - \delta) \ln(2)} \exp \left( - \frac{1}{4 \ln \left( \frac{1}{\alpha_\varepsilon} \right)} \right). \end{aligned}$$

Then taking the asymptotic.

$$(5.10) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{\theta \in \Sigma - \Theta_{\varepsilon,\delta}} F_\varepsilon(\theta) \leq \frac{\lambda^2}{2\delta(1 - \delta) \ln(2)}.$$

According to (5.10) and (5.8), it provides that

$$(5.11) \quad \limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(1,2)} \leq L^2 + \inf_{1 > \delta > 0} \frac{\lambda^2}{2\delta(1 - \delta) \ln(2)} = L^2 + \frac{2\lambda^2}{\ln(2)}.$$

The results comes from (5.5), (5.6) and (5.11).

**5.2.2. Lower bound.** We will check the optimality of  $(\rho_\varepsilon^*)$  defined in (3.15), considering an arbitrary RNF  $\tilde{\rho}_\varepsilon = \{a_\varepsilon(\tilde{\rho}), \varphi_\varepsilon(\Sigma)\} \in \Omega$  such that

$$\frac{a_\varepsilon(\tilde{\rho})}{a_\varepsilon} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Let us denoted by  $B_\varepsilon$  the event corresponding to the acceptance of additivity w.r.t.  $\tilde{\rho}_\varepsilon$ , then

$$B_\varepsilon = \{\tilde{\rho}_\varepsilon = a_\varepsilon(\tilde{\rho})\}.$$

Let

$$\begin{aligned} \psi_\varepsilon &= \frac{C}{2^{\frac{d+1}{4}}} \left( \varepsilon^2 \sqrt{\ln(1/\alpha_\varepsilon)} \right)^{(2\beta+d)/(4\beta+d)}, \\ N_{1,\varepsilon} &= C^{-\frac{4}{d}} N_{0,\varepsilon} = 2C^{-\frac{4}{d}} \left( \varepsilon^2 \sqrt{\ln(1/\alpha_\varepsilon)} \right)^{-2/(4\beta+d)} \end{aligned}$$

when the constant  $C$  is given by

$$C = \left[ \frac{L^2}{2^{2\beta+(d-1)/2}(d+1)} \right]^{-\frac{d}{8\beta+2d}}.$$

For a fixed  $\varepsilon > 0$ , we define the subset of  $\mathbb{N}^d$

$$\Delta_{1,\varepsilon} = \{k \in \mathbb{N}^d : k_j < N_{1,\varepsilon} \quad \forall j = 1\},$$

Put  $V = \{-1, 1\}^{|\Delta_{1,\varepsilon}|}$ , then we consider the family  $\mathcal{U}_\varepsilon$  of sequences  $\theta(v), v \in V$  indexed by  $\mathbb{N}^d$  such that

$$\forall k \in \mathbb{N}^d, \quad \theta(v)_k = \psi_\varepsilon v_k \mathbf{1}_{k \in \Delta_{1,\varepsilon}}.$$

In a first time, let's check that  $\mathcal{U}_\varepsilon$  is included in  $\Sigma$ . We just need to remark that for any  $v \in V$ , we have according to the choice of the constant  $C$  :

$$\begin{aligned} \sum_{k \in \mathbb{N}^d} \theta(v)_k^2 \left( 1 + \sum_{i=1}^d k_i^{2\beta} \right) &= \psi_\varepsilon^2 \sum_{k \in \Delta_{1,\varepsilon}} \left( 1 + \sum_{i=1}^d k_i^{2\beta} \right) \\ &\leq (d+1) \psi_\varepsilon^2 N_{1,\varepsilon}^{2\beta+d} = L^2. \end{aligned}$$

In the following we denote by  $P_0$  the probability measure w.r.t. observations if the true function is  $f(\cdot) = 0$ . If  $f(\cdot) = \sum_{k \in \mathbb{N}^d} \theta(v)_k \phi_k(\cdot)$ , then we note the law of observations by  $P_v$ . For one given  $k \in \Delta_{1,\varepsilon}$  let us define the following subsets of  $V$  :

$$V_k^{(1)} = \{v \in V : v_k = 1\}, \quad V_k^{(-1)} = \{v \in V : v_k = -1\},$$

$$V_k^{(0)} = \{v = (v_l)_{l \in \Delta_{1,\varepsilon}} : \text{if } l \neq k, v_l \in \{-1, 1\} \text{ and } v_k = 0\}.$$

Thus, for any  $v \in V$ , we define the sequence  $v^{(k)} = (v_l^{(k)})_{l \in \Delta_{1,\varepsilon}} \in V_k^{(0)}$  by :

$$v_l^{(k)} = \begin{cases} v_l, & \text{if } l \neq k, \\ 0, & \text{if } l = k. \end{cases}$$

We will prove that there exists  $p_0 > 0$  such that for any  $k \in \Delta_{1,\varepsilon}$ ,

$$(5.12) \quad \forall k \in \Delta_{1,\varepsilon}, \quad \frac{1}{2^{N_{1,\varepsilon}^d - 1}} \sum_{v \in V_k^{(0)}} P_v(B_\varepsilon) \geq p_0.$$

Let  $\widehat{\theta}_\varepsilon$  an arbitrary estimator of  $\theta$ . We aim to give a lower bound for its risk (2.1) normalized by  $(\widetilde{\rho}_\varepsilon)_\varepsilon$ , focusing on the specific estimation error on each element of  $U_n$ .

$$\begin{aligned} R_\varepsilon(\widehat{\theta}_\varepsilon, \widetilde{\rho}_\varepsilon, \Sigma) &\geq \sup_{\theta \in \Sigma} E_\theta \left\{ \widetilde{\rho}_\varepsilon^{-2} \|\widehat{\theta}_\varepsilon - \theta\|^2 1_{B_\varepsilon} \right\} \\ &\geq \sup_{\theta(v): v \in V} E_\theta \left\{ a_\varepsilon(\widetilde{\rho})^{-2} \|\widehat{\theta}_\varepsilon - \theta\|^2 1_{B_\varepsilon} \right\} \\ &\geq \frac{a_\varepsilon(\widetilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{v \in V} E_v \left\{ \|\widehat{\theta}_\varepsilon - \theta\|^2 1_{B_\varepsilon} \right\} \\ &\geq \frac{a_\varepsilon(\widetilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{v \in V} \sum_{k \in \Delta_{1,\varepsilon}} E_v \left\{ (\widehat{\theta}_{\varepsilon,k} - \theta(v)_k)^2 1_{B_\varepsilon} \right\} \\ (5.13) \quad &\geq \frac{a_\varepsilon(\widetilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{k \in \Delta_{1,\varepsilon}} \left( \sum_{v \in V_k^{(1)}} E_v \left\{ (\widehat{\theta}_{\varepsilon,k} - \psi_\varepsilon)^2 1_{B_\varepsilon} \right\} + \sum_{v \in V_k^{(-1)}} E_v \left\{ (\widehat{\theta}_{\varepsilon,k} + \psi_\varepsilon)^2 1_{B_\varepsilon} \right\} \right). \end{aligned}$$

Let us note that according to the Girsanov theorem, under  $P_{v^{(k)}}$

$$\begin{aligned} \forall v \in V_k^{(1)}, \quad Z_k^{(1)} &= \frac{dP_v}{dP_{v^{(k)}}} = \exp \left\{ \psi_\varepsilon \varepsilon^{-1} \zeta_k - \frac{\varepsilon^{-2} \psi_\varepsilon^2}{2} \right\}, \\ \forall v \in V_k^{(-1)}, \quad Z_k^{(-1)} &= \frac{dP_v}{dP_{v^{(k)}}} = \exp \left\{ -\psi_\varepsilon \varepsilon^{-1} \zeta_k - \frac{\varepsilon^{-2} \psi_\varepsilon^2}{2} \right\} \end{aligned}$$

when  $(\zeta_k)_k$  is standard Gaussian sample. As  $\psi_\varepsilon \varepsilon^{-1} \rightarrow 0$  when  $\varepsilon \rightarrow 0$ , these two random variables tend in probability to 1. Thus, it exists a real  $\delta > 0$  independent from index  $k$  such that for small enough  $\varepsilon$  (typically  $\varepsilon \leq \varepsilon_0$ ), one have

$$(5.14) \quad P_{v^{(k)}} \left( \left\{ Z_k^{(1)} < 1 - \delta \right\} \cup \left\{ Z_k^{(-1)} < 1 - \delta \right\} \right) \leq \delta, \quad \forall k \in \Delta_{1,\varepsilon}.$$

In the following we denoted by  $D_k = \left\{ Z_k^{(1)} \geq 1 - \delta \right\} \cap \left\{ Z_k^{(-1)} \geq 1 - \delta \right\}$ . Then,

$$\begin{aligned} R_\varepsilon(\hat{\theta}_\varepsilon, \tilde{\rho}_\varepsilon, \Sigma) &\geq \frac{a_\varepsilon(\tilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{k \in \Delta_{1,\varepsilon}} \sum_{v \in V_k^{(1)}} E_{v^{(k)}} \left\{ Z_k^{(1)} (\hat{\theta}_{\varepsilon,k} - \psi_\varepsilon)^2 1_{B_\varepsilon} \right\} \\ &+ \frac{a_\varepsilon(\tilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{k \in \Delta_{1,\varepsilon}} \sum_{v \in V_k^{(-1)}} E_{v^{(k)}} \left\{ Z_k^{(-1)} (\hat{\theta}_{\varepsilon,k} + \psi_\varepsilon)^2 1_{B_\varepsilon} \right\} \\ &\geq \frac{a_\varepsilon(\tilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{k \in \Delta_{1,\varepsilon}} \sum_{v \in V_k^{(0)}} E_v \left\{ \left[ Z_k^{(1)} (\hat{\theta}_{\varepsilon,k} - \psi_\varepsilon)^2 + Z_k^{(-1)} (\hat{\theta}_{\varepsilon,k} + \psi_\varepsilon)^2 \right] 1_{B_\varepsilon \cap D_k} \right\}. \end{aligned}$$

Using (5.14) and that  $(a - b)^2 + (a + b)^2 \geq 2b^2$ , one get

$$\begin{aligned} R_\varepsilon(\hat{\theta}_\varepsilon, \tilde{\rho}_\varepsilon, \Sigma) &\geq \frac{2(1 - \delta) \psi_\varepsilon^2 a_\varepsilon(\tilde{\rho})^{-2}}{2^{|\Delta_{1,\varepsilon}|}} \sum_{k \in \Delta_{1,\varepsilon}} \sum_{v \in V_k^{(0)}} P_v \{ B_\varepsilon \cap D_k \} \\ &\geq 2(1 - \delta) \psi_\varepsilon^2 a_\varepsilon(\tilde{\rho})^{-2} \sum_{k \in \Delta_{1,\varepsilon}} \frac{1}{2^{|\Delta_{1,\varepsilon}|}} \sum_{v \in V_k^{(0)}} [P_v(B_\varepsilon) - P_v(D_k^c)]. \end{aligned}$$

According to (5.12) and (5.14), as soon as we select  $\delta$  sufficiently small, that is for instance  $\delta < (\frac{p_0}{2} \wedge \frac{1}{2})$  it follows that

$$R_\varepsilon(\hat{\theta}_\varepsilon, \tilde{\rho}_\varepsilon) \geq \frac{p_0}{2} a_\varepsilon(\tilde{\rho})^{-2} \psi_\varepsilon^2 |\Delta_{1,\varepsilon}|.$$

From  $\psi_\varepsilon^2 |\Delta_{1,\varepsilon}| \geq C^{-2} a_\varepsilon^2$  and since  $\frac{a_\varepsilon(\tilde{\rho})}{a_\varepsilon} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , one deduces that

$$(5.15) \quad \lim_{\varepsilon \rightarrow 0} R_\varepsilon(\hat{\theta}_\varepsilon, \tilde{\rho}_\varepsilon, \Sigma) = +\infty.$$

We complete the proof by proving (5.12). For a given  $k \in \Delta_{1,\varepsilon}$ , let us note  $m_\varepsilon = |V_k^{(0)}| = N_{1,\varepsilon}^d - 1$  and consider

$$\begin{aligned} J_\varepsilon &= J_\varepsilon(k) = \frac{1}{2^{m_\varepsilon}} \sum_{v \in V_k^{(0)}} P_v(B_\varepsilon), \\ Z_\varepsilon &= Z_\varepsilon(k) = \frac{1}{2^{m_\varepsilon}} \sum_{v \in V_k^{(0)}} \frac{dP_v}{dP_0}(Y). \end{aligned}$$

Now let us write probabilities under the measure induced by  $v = 0$ , we get  $J_\varepsilon = E_0\{Z_\varepsilon 1_{B_\varepsilon}\}$ , thus as  $0 \in \Sigma_0$ , for any  $\delta > 0$ , and for  $n$  large enough :

$$\alpha_\varepsilon^{-1} P_0(B_\varepsilon^c) \leq 1 + \delta.$$

Then for any  $c > 0$ ,

$$\begin{aligned}
J_\varepsilon &\geq E_0 \{ Z_\varepsilon 1_{B_\varepsilon} + c\alpha_\varepsilon^{-1} 1_{B_\varepsilon^c} \} - c(1 + \delta) \\
&\geq E_0 \left\{ \inf_{F_\varepsilon} [Z_\varepsilon 1_{F_\varepsilon} + c\alpha_\varepsilon^{-1} 1_{F_\varepsilon^c} - c(1 + \delta)] \right\} \\
&\geq E_0 \{ Z_\varepsilon 1_{Z_\varepsilon < c\alpha_\varepsilon^{-1}} \} - c(1 + \delta) \\
&\geq \frac{1}{2^{m_\varepsilon}} \sum_{v \in V_k^{(0)}} P_v(Z_\varepsilon < c\alpha_\varepsilon^{-1}) - c(1 + \delta) \\
&\geq 1 - c(1 + \delta) - \frac{1}{2^{m_\varepsilon}} \sum_{v \in V_k^{(0)}} P_v(Z_\varepsilon \geq c\alpha_\varepsilon^{-1}) \\
&\geq 1 - c(1 + \delta) - \frac{\alpha_\varepsilon}{c2^{m_\varepsilon}} \sum_{v \in V_k^{(0)}} E_v(Z_\varepsilon) \\
(5.16) \quad &\geq 1 - c(1 + \delta) - c^{-1}\alpha_\varepsilon E_0(Z_\varepsilon^2).
\end{aligned}$$

The distribution of  $Z_\varepsilon$  under  $P_0$  is given by Girsanov theorem,

$$Z_\varepsilon = \frac{1}{2^{m_\varepsilon}} \exp\left(-\frac{m_\varepsilon \psi_\varepsilon^2}{2\varepsilon^2}\right) \sum_{v \in V_k^{(0)}} \exp\left(\psi_\varepsilon \varepsilon^{-1} \sum_j v_j \zeta_j\right).$$

Then, its second moment is :

$$\begin{aligned}
Z_\varepsilon^2 &= \frac{\exp(-m_\varepsilon \psi_\varepsilon^2 \varepsilon^{-2})}{2^{2m_\varepsilon}} \sum_{v^1, v^2 \in V_k^{(0)}} \exp\left(\psi_\varepsilon \varepsilon^{-1} \sum_j (v_j^1 + v_j^2) \zeta_j\right) \\
&= \frac{\exp(-m_\varepsilon \psi_\varepsilon^2 \varepsilon^{-2})}{2^{2m_\varepsilon}} \sum_{v^1, v^2 \in V_k^{(0)}} \exp\left(2\psi_\varepsilon \varepsilon^{-1} \sum_j v_j^1 1_{v_j^1 = v_j^2} \zeta_j\right) \\
(5.17) \quad &= \frac{\exp(-m_\varepsilon \psi_\varepsilon^2 \varepsilon^{-2})}{2^{2m_\varepsilon}} \sum_{v^1 \in V_k^{(0)}} \sum_{p=0}^{m_\varepsilon} \sum_{J: |J|=p} \exp\left(2\psi_\varepsilon \varepsilon^{-1} \sum_{j \in J} v_j^1 \zeta_j\right).
\end{aligned}$$

However as  $(\zeta_j)_j$  is supposed to be a family of i.i.d standard Gaussian variables, we get that  $E(e^{a\zeta}) = e^{\frac{a^2}{2}}$ . Finally (5.17) becomes :

$$\begin{aligned}
E_0(Z_\varepsilon^2) &= \frac{\exp(-m_\varepsilon \psi_\varepsilon^2 \varepsilon^{-2})}{2^{2m_\varepsilon}} \sum_{v^1 \in V_k^{(0)}} \sum_{p=0}^{m_\varepsilon} \binom{m_\varepsilon}{p} e^{2p\psi_\varepsilon^2 \varepsilon^{-2}} \\
&= \frac{\exp(-m_\varepsilon \psi_\varepsilon^2 \varepsilon^{-2})}{2^{m_\varepsilon}} \left(1 + e^{2\psi_\varepsilon^2 \varepsilon^{-2}}\right)^{m_\varepsilon} \\
(5.18) \quad &= \left(\frac{e^{\psi_\varepsilon^2 \varepsilon^{-2}}}{2} + \frac{e^{-\psi_\varepsilon^2 \varepsilon^{-2}}}{2}\right)^{m_\varepsilon} = ch(\psi_\varepsilon^2 \varepsilon^{-2})^{m_\varepsilon}.
\end{aligned}$$

It can be underlined that the expectation does not depend on  $k \in \Delta_{1,\varepsilon}$ . According to the choice of parameter, this term is asymptotically controlled by  $\alpha_\varepsilon^{-1/2}$ . Indeed, using that  $\forall x \in \mathbb{R}$ ,  $ch(x) \leq e^{x^2}$  one get

$$(5.19) \quad E_0(Z_\varepsilon^2) \leq \exp(N_{1,\varepsilon}^d \psi_\varepsilon^4 \varepsilon^{-4})(1 + o(1)) = \alpha_\varepsilon^{-1/2}(1 + o(1)).$$

Supposing for instance that  $\alpha_\varepsilon \leq 1/20$  and after optimization w.r.t.  $c > 0$ , we deduce from (5.16) and (5.19) that

$$J_\varepsilon \geq 1 - \frac{2}{\sqrt{2\sqrt{5}}}(1 + o(1)).$$

This ends the proof of theorem 3.1.

**5.3. Proof of proposition 4.1.** Assume that  $f \in \Sigma_0$ , we have :

$$P_f(\rho^* \neq a_n) = P_f(A_n^c) = P_f(T_n > \lambda a_n^2),$$

when according to (3.10) and (4.18) ,

$$\begin{aligned} T_n &= \sum_{k \in \Lambda_n} \theta_k^2(\mathcal{J}_{n-1}(f)) + \sum_{k \in \Lambda_n} (\eta_k^2 - \sigma_k^2) = \sum_{k \in \Lambda_n} b_k^2(f) + \sum_{k \in \Lambda_n} (\eta_k^2 - \sigma_k^2) + 2 \sum_{k \in \Lambda_n} \eta_k b_k(f) \\ &\leq \sum_{k \in \mathbb{N}^d} b_k^2(f) + \sum_{k \in \Lambda_n} (\eta_k^2 - \sigma_k^2) + 2 \sum_{k \in \Lambda_n} \eta_k b_k(f) \\ (5.20) \quad &\leq \|f - \mathcal{J}_{n-1}f\|_{L_2}^2 + \mathcal{S}(\Lambda_n) + Z_{\Lambda_n} \end{aligned}$$

where  $Z_{\Lambda_n} \sim \mathcal{N}(0, 4 \sum_{k \in \Lambda_n} \sigma_k^2 b_k^2(f))$  and when for any  $A \subset \mathbb{N}^d$ , we set  $\mathcal{S}(A) = \sum_{k \in A} (\eta_k^2 - \sigma_k^2)$ . Thus, under assumption that (4.16) holds,

$$(5.21) \quad P_f(\rho_n^* \neq a_n) \leq P_f\left(c_2 n^{-\frac{2m}{d}} + \mathcal{S}(\Lambda_n) + Z_{\Lambda_n} > \lambda a_n^2\right).$$

From  $m > \frac{d}{4}$  one get that  $a_n^{-2} n^{-\frac{2m}{d}} \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover usual large deviations of Gaussian distribution lead to

$$(5.22) \quad P_f\left(Z_{\Lambda_n} > \frac{\lambda a_n^2}{3}\right) \leq e^{-\frac{\mu(\lambda) a_n^4}{\sum_{k \in \Lambda_n} \sigma_k^2 b_k^2(f)}}.$$

where  $\mu(\lambda) > 0$ . Sincet  $(\eta_k)_{k \in \Lambda_n}$  is a Gaussian vector with non correlated coordinates,  $\eta_k$  are independant variables. Condition 2 that proves that  $0 < \lambda_{i,n} \leq c_1 n^{-1}$  yields

$$\forall k \in \{1, \dots, n-1\}^d, \quad \sigma_k^2 = \sum_{\mathbf{i}} \lambda_{\mathbf{i},n}^2 \Phi_k^2(x_{\mathbf{i}}) \leq \frac{c_1}{n} \sum_{\mathbf{i}} \lambda_{\mathbf{i},n} \Phi_k^2(x_{\mathbf{i}})$$

But according to optimality of quadrature formula (4.12), one deduces

$$(5.23) \quad \forall k \in \{1, \dots, n-1\}^d, \quad \sigma_k^2 \leq \frac{c_1}{n} \|\Phi_k\|^2 = \frac{c_1}{n}.$$

Add that with (4.16) for any  $f \in \Sigma$ ,

$$(5.24) \quad \sum_{k \in \Lambda_n} b_k^2(f) \leq \sum_{k \in \mathbb{N}^d} b_k^2(f) \leq \|f - \mathcal{J}_{n-1}f\|_{L_2}^2 \leq c n^{-\frac{2m}{d}}.$$

So using both (5.23) and (5.24), it follows that  $\sum_{k \in \Lambda_n} \sigma_k^2 b_k^2(f) \ll a_n^4$ , and then there exists  $\kappa > 0$  only depending on  $m$  and  $d$  such that

$$(5.25) \quad \forall f \in \Sigma, \quad P_f\left(Z_{\Lambda_n} > \frac{\lambda a_n^2}{3}\right) \leq e^{-n^\kappa}$$

Then, from (5.21), (5.25) and for large enough  $n \in \mathbb{N}$ ,

$$(5.26) \quad P_f(\rho_n^* \neq a_n) \leq P_f\left(\mathcal{S}(\Lambda_n) > \frac{\lambda a_n^2}{3}\right) + e^{-n^\kappa}.$$

To complete the proof, it is sufficient to use similar large deviation that in proof of lemma 3.1 and above remark. Result comes in the same way.

#### 5.4. Proof of theorem 4.1.

$$\begin{aligned} R_n^{(r)}(f_n^*, \rho_n^*) &= \sup_{f \in \Sigma} E_f \left\{ (\rho_n^*)^{-2} \|f_n^* - f\|^2 \right\} \\ &\leq 2 \left\{ a_n^{-2} \sup_{f \in \Sigma} \|f - \mathcal{J}_{n-1}f\|^2 + \sup_{f \in \Sigma} E_f \left\{ (\rho_n^*)^{-2} \|\mathcal{J}_{n-1}f - f_n^*\|_\tau^2 \right\} \right\}. \end{aligned}$$

But according to (4.16) one obtains for any  $f \in \Sigma$ ,

$$\begin{aligned} R_n^{(r)}(f_n^*, \rho_n^*) &\leq 2 \left\{ c^2 a_n^{-2} M_n^{-2m} + \sup_{f \in \Sigma} E_f \left\{ (\rho_n^*)^{-2} \|\mathcal{J}_{n-1}f - f_n^*\|^2 \right\} \right\} \\ &\leq 2 \left\{ c^2 a_n^{-2} n^{-\frac{2m}{d}} + \sup_{f \in \Sigma} R_n^{(1)}(f) + \sup_{f \in \Sigma} R_n^{(2)}(f) \right\}. \end{aligned}$$

when we put

$$\begin{aligned} R_n^{(1)}(f) &= E_f \left\{ a_n^{-2} \|\mathcal{J}_{n-1}f - \widehat{f}_n^{(0)}\|^2 \mathbf{1}_{A_n} \right\} \\ R_n^{(2)}(f) &= E_f \left\{ \varphi_n(\Sigma)^{-2} \|\mathcal{J}_{n-1}f - \widehat{f}_n\|^2 \mathbf{1}_{A_n^c} \right\}. \end{aligned}$$

Once  $m > \frac{d}{4}$ , then  $a_n^{-2} n^{-\frac{2m}{d}} \rightarrow 0$  as  $n \rightarrow \infty$ . Then for any  $f \in \Sigma$ ,

$$\begin{aligned} R_n^{(1)}(f) &= a_n^{-2} \left( \sum_{k \notin I_n} \theta_k^2(\mathcal{J}_{n-1}(f)) \right) P_f(A_n) + a_n^{-2} E_f \left( \mathbf{1}_{A_n} \sum_{k \in I_n} \eta_k^2 \right) \\ (5.27) \quad R_n^{(1)}(f) &\leq a_n^{-2} \left( \sum_{k \notin I_n} \theta_k^2(f) + \sum_{k \notin I_n} [b_k^2(f) + 2\theta_k(f)b_k(f)] \right) P_f(A_n) + a_n^{-2} \sum_{k \in I_n} \sigma_k^2. \end{aligned}$$

From (5.23), it follows that

$$(5.28) \quad a_n^{-2} \sum_{k \in I_n} \sigma_k^2 \longrightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Denoted by

$$\begin{aligned} H_n(f) &= \sum_{k \in \Lambda_n} \theta_k^2(f) \quad \text{and} \quad K_n(f) = \sum_{k \notin \Delta_n} \theta_k^2(f), \\ \mathcal{V}(I_n) &= \sum_{k \notin I_n} [b_k^2(f) + 2\theta_k(f)b_k(f)], \\ \mathcal{V}(\Lambda_n) &= \sum_{k \in \Lambda_n} [b_k^2(f) + 2\theta_k(f)b_k(f)]. \end{aligned}$$

Let us also define

$$\begin{aligned} \underline{H}_n(f) &= H_n(f) + \mathcal{V}(I_n), \\ \widetilde{H}_n(f) &= H_n(f) + \mathcal{V}(\Lambda_n) = \underline{H}_n(f) + \mathcal{V}(\Lambda_n) - \mathcal{V}(I_n). \end{aligned}$$

Therefore, using (5.28) and that for any  $f \in \Sigma$ ,  $a_n^{-2} K_n(f) \leq L^2$  we derive from (5.27) that

$$\begin{aligned} R_n^{(1)}(f) &\leq a_n^{-2} (\underline{H}_n(f) + K_n(f)) P_f(A_n) + o_n(1) \\ (5.29) \quad &\leq L^2 + a_n^{-2} \underline{H}_n(f) P_f(T_n < \lambda a_n^2) + o_n(1). \end{aligned}$$

Introduce

$$(5.30) \quad \Theta_{n,\delta} = \left\{ f \in \Sigma : \underline{H}_n(f) \leq \frac{1+\delta}{1-\delta} \lambda^2 a_n^2 \right\}.$$

Now put

$$R_n^{(1,1)} = \sup_{f \in \Theta_{n,\delta}} R_n^{(1)}(f), \quad \text{and} \quad R_n^{(1,2)} = \sup_{f \in \Sigma - \Theta_{n,\delta}} R_n^{(1)}(f).$$

First from (5.29), one derive the asymptotic

$$(5.31) \quad \limsup_n R_n^{(1,1)} \leq L^2 + \frac{1+\delta}{1-\delta} \lambda^2,$$

if we use same decomposition of  $\Sigma - \Theta_{n,\delta} = \bigcup_{x>1+\delta} \Theta(x)$  where

$$\Theta(x) = \left\{ f \in \Sigma : \underline{H}_n(f) = \frac{x\lambda^2 a_n^2}{1-\delta} \right\}.$$

Assume that  $f \in \Theta(x)$ , from (5.29) one have

$$(5.32) \quad R_n^{(1)}(f) \leq L^2 + \frac{x\lambda^2}{1-\delta} P_f(T_n < \lambda a_n^2) + o_n(1).$$

Let us decompose

$$(5.33) \quad T_n = \tilde{H}_n(f) + \zeta_n(f) + \mathcal{S}(\Lambda_n),$$

where

$$\zeta_n(f) = 2 \sum_{k \in \Lambda_n} \eta_k(\theta_k(f) + b_k(f))$$

is a Gaussian random variable with zero mean and variance

$$(5.34) \quad \text{Var}(\zeta_n(f)) = 4 \sum_{k \in \Lambda_n} \sigma_k^2(\theta_k(f) + b_k(f))^2 \leq \frac{4c_1}{n} \tilde{H}_n(f).$$

Then for any  $f \in \Sigma$ , there exists  $\gamma > 0$  such that

$$(5.35) \quad P_f(|\zeta_n(f)| > \delta \tilde{H}_n(f)) \leq 2 \exp \left\{ -\frac{\delta^2 n \tilde{H}_n(f)}{2} \right\} \leq \exp \{-n^\gamma\}.$$

Consequently for any  $f \in \Theta(x)$ ,

$$(5.36) \quad \begin{aligned} P_f(T_n \leq \lambda^2 a_n^2) &\leq P_f(\tilde{H}_n(f)(1-\delta) + \mathcal{S}(\Lambda_n) \leq \lambda^2 a_n^2) + \exp\{-n^\gamma\}(1+o(1)) \\ &\leq P_f(\underline{H}_n(f)(1-\delta) + \mathcal{V}(\Lambda_n) - \mathcal{V}(I_n) + \mathcal{S}(\Lambda_n) \leq \lambda^2 a_n^2) + \exp\{-n^\gamma\}(1+o(1)). \end{aligned}$$

Moreover according to (4.16), the contribution of the interpolation step in bias can be neglected. Indeed, when  $M = M_n$

$$\begin{aligned} |\mathcal{V}(\Lambda_n) - \mathcal{V}(I_n)| &= \left| \sum_{k \notin \Delta_n} [b_k^2(f) + 2\theta_k(f)b_k(f)] \right| \\ &\leq \sum_{k \notin \mathbb{N}^d} b_k^2(f) + 2 \sqrt{\sum_{k \notin \Delta_n} \theta_k^2(f)} \sqrt{\sum_{k \notin \Delta_n} b_k^2(f)} \\ &\leq \|f - \mathcal{J}_{n-1}f\|_{L_2}^2 + 2\|f - \mathcal{J}_{n-1}f\| \sqrt{K_n(f)} \\ &\leq c'^2 L'^2 n^{-\frac{2m}{d}} + c' L' L n^{-\frac{m}{d}} \sqrt{K_n(f)}. \end{aligned}$$

Then from  $a_n^{-2} n^{-\frac{2m}{d}} \rightarrow 0$ , and that  $\forall f \in \Sigma$ ,  $a_n^{-2} K_n(f) \leq L^2$ , one derive that  $|\mathcal{V}(\Lambda_n) - \mathcal{V}(I_n)| \ll a_n^2$ . In particular for large enough  $n$ , one have

$$(5.37) \quad \forall f \in \Sigma, \quad |\mathcal{V}(\Lambda_n) - \mathcal{V}(I_n)| \leq \frac{\lambda^2 a_n^2}{2}.$$



Then (5.36) gives

$$\begin{aligned}
 P_f(T_n \leq \lambda^2 a_n^2) &\leq P_f\left(\underline{H}_n(f)(1 - \delta) + \mathcal{S}(\Lambda_n) \leq \frac{\lambda^2 a_n^2}{2}\right) + \exp\{-n^\gamma\}(1 + o(1)) \\
 (5.38) \qquad \qquad \qquad &\leq P_f\left(\mathcal{S}(\Lambda_n) \leq -\frac{(x-1)\lambda^2 a_n^2}{2}\right) + \exp\{-n^\gamma\}(1 + o(1)).
 \end{aligned}$$

It follows from (5.29) and (5.38) that

$$(5.39) \qquad \sup_{f \in \Theta(x)} R_n^{(1)}(f) \leq L^2 + \frac{x\lambda^2}{2(1-\delta)} P_f\left(\mathcal{S}(\Lambda_n) \leq -\frac{(x-1)\lambda^2 a_n^2}{2}\right) + o_n(1).$$

## REFERENCES

- [1] Barron A., Birgé L. and Massart P. (1999) *Risk bounds for model selection via penalization* .Probab. Theory Related Fields **113** 301-413
- [2] Bernardi C., Maday Y. (1992) *Aproximations spectrales de problèmes aux limites elliptiques* . Springer-Verlag
- [3] Breiman L., Friedman J.H., (1985) *Estimating optimal transformations for multiple regression and correlations (with discussion)* J.of the American Statist. Assoc. **80** 580-619
- [4] Brown L., Low M. (1996) *a constrained risk inequality with applications to nonparametric estimation* Annals of Stat. **24** 2524-2535
- [5] Ciarlet P.G. (1982) *Introduction à l'analyse numérique matricielle et à l'optimisation* Masson
- [6] Chen H. (1991) *Estimation of a projection-pursuit type regression model* Annals of Stats. **19** 142-157
- [7] Donoho D.L., Johnstone I.M. (1994) *Ideal spatial adaptation by wavelet shrinkage*.Biometrika **81** 425-455
- [8] Donoho D.L., Johnstone I.M., Kerkycharian G., Picard D. (1995) *Wavelet shrinkage : Asymptopia*. J.Roy.Statist.Soc.Ser.**57** 301-369
- [9] Efremovich S., Pinsker M.S., (1984) *Adaptive algorithms for nonparametric filtering* Automat.Remote Control **11** 54-60
- [10] Golubev G.K. (1990) *Quasilinear estimates of a signal in  $L_2$*  Problems Inform. Transmission **26** 15-20
- [11] Golubev G.K. (1992) *Asymptotically minimax estimation of a regression function on an additive model* Problems Inform. Transmission **28** 101-112
- [12] Hardle W. and Marron M. (1985) *Optimal bandwidth selection in nonparametric regression function estimation* Ann. Statist.**3** 1465-1481
- [13] Hardle W., Tsybakov A.B., (1995) *Additive nonparametric regression on principal components* J. of Nonparametric Stats. **5** 157-184
- [14] Hall P., Kerkycharian G., Picard D. (1998) *Block threshold rules for curve estimation using kernel and wavelet methods* Ann.Statist. **26** 922-942
- [15] Hall P., Kerkycharian G., Picard D. (1999) *On the minimax optimality of block thresholding wavelet estimators* Statist. Sinica **9** 33-49
- [16] Ibragimov I.A., Hasminskii R.Z., (1980) *On non parametric estimation of regression* Soviet Math. Dokl.
- [17] Lepskii, O. (1990) *A problem of adaptive estimation in gaussian white noise* Theory Prob.Appli.**35** 454-466.
- [18] Lepskii, O. (1991) *Asymptotic minimax adaptive estimation. 1. Upper bounds. Optimally adaptive estimates* Theory Prob.Appli. **36** 682-697
- [19] Lepskii, O. (1992) *Asymptotic minimax adaptive estimation. 2. Statistical models without optimal adaptation. Adaptive estimators* Theory Prob.Appli. **37**433-448.
- [20] Lepskii, O. (1992) *On problems of adaptive estimation in white gaussian noise* Soviet Mathematics **12** 87-106.
- [21] Lepskii, O. *How to improve the accuracy of estimaton* .Math.Meth.of Stat. (1999)
- [22] Lepskii O., Hoffmann M. *Random rates in anisotropic regression* . Annals of Stat. (2002) **30** 325-396
- [23] Lepskii O., Spokoiny V.G. (1995) *Local adaptation to inhomogenous smoothness : resolution level*. Math.Methods Statist. **4** 239-258
- [24] Lepskii O., Spokoiny V.G. (1997) *Optimal pointwise adaptive methods in non parametric estimation* Ann.Statist. **25** 2512-2546
- [25] Lepskii O., Mammen E., Spokoiny V.G. (1997) *Optimal spatial adaptation to inhomogenous smoothness / an approach based on kernel estimates with variable bandwidth selectors* Ann.Statist. **25** 929-947
- [26] Linton O., Nielsen J.P. (1995) *A kernel method of estimating structured nonparametric regression based on marginal integration* Biometrika **82** 93-101
- [27] Mammen E., Linton O., Nielsen J.P. (1999) *The existence and asymptotic properties of a backfitting projection algorithm under weak conditions* Annals of Stats **27** 1443-1490
- [28] Poliak B., Tsybakov A.B. (1990) *Asymptotic optimality of the  $C_p$  test in the projection estimation of a regression* Theory Prob.Appl. **35** 293-306
- [29] Severance-Lossin E., Sperlich S. (1999) *Estimation of derivatives for additive separable models* Statistics **33** 241-265
- [30] Stone C.J. *The dimensionality reduction principle for generalized additive models* Annals of Stat. **14** 590-606

- [31] Szego G. (1978) *Orthogonal Polynomials*, Colloquium Publication AMS, Providence
- [32] Yode A.F.E. (2004) *Estimation de la densité de probabilité multidimensionnel*